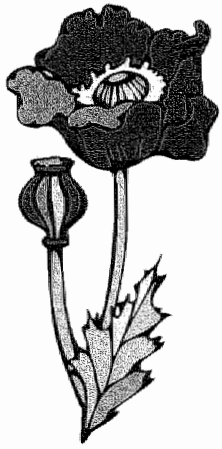
In-training assessment in a work-based postgraduate medical education context

In-training assessment

in a work-based postgraduate medical education context



ISBN 90-5278421-3
Datawyse | Universitaire Pers Maastricht, the Netherlands

In-training assessment

in a work-based postgraduate medical education context

PROEFSCHRIFT

ter verkrijging van de graad van doctor
aan de Universiteit Maastricht,
op gezag van de Rector Magnificus,
Prof. G.P.M.F. Mols
volgens het besluit van het College van Decanen,
in het openbaar te verdedigen
op vrijdag 2 juli 2004 om 16.00 uur

door

Charlotte Ringsted

UNIVERSITAIRE PERS MAASTRICHT

PROMOTORES

Prof. dr. A.J.J.A. Scherpbier
Prof. dr. C.P.M. van der Vleuten

BEOORDELINGSCOMMISSIE

Prof. dr. M.von Meyenfeldt (voorzitter)
Ass. Prof. dr B. Eika (Aarhus University, Denmark)
Prof. dr. M.van Kleef
Dr. Y van Leeuwen
Prof. dr. E. Petrusa (Duke University, South Carolina, USA)

The research studies of this thesis were supported by the Copenhagen Hospital Corporation Postgraduate Medical Institute and in part by grants from the Ministry of Health, Denmark and from the Danish Society of Anaesthesiology and Intensive Care Medicine.

To Camilla, Thomas, and Tinne– who are my joy and inspiration
To my father and mother – who gave me love, strength, and courage

Table of contents

Introduction

Postgraduate specialist education is built on top of a long undergraduate medical education and in some countries an additional internship. This makes it the longest formal education seen in any profession with a huge investment of human and financial resources. For accountability reasons alone there is cause to make this education as effective and efficient as possible and to provide evidence of the outcome.[1] Some unique conditions of postgraduate education are additional significant reasons to assure efficacy and quality of the programmes. These conditions pose a major challenge to the educators who are responsible for choosing assessment strategies with sufficient feasibility and educational impact. Postgraduate education is primarily work-based and the learning environment is not a teaching institution but a highly complex and busy working context of which the primary role is health service. The teachers are clinicians and they do not necessarily have strong incentives or protected time for their role as educators.[2] Both trainees and trainers often perceive education in conflict with the service demands and hence efficient and feasible strategies of education and assessment are needed.[2,3] Postgraduate trainees constitute a substantial part of the medical work force[3] and trainees frequently work unsupervised.[4,5] This means that quality of care and patient safety rely on strategies that ensure trainees' competence along the way during their training as well as by the end of training.[4,5,6]

Society requires responsibility from the medical profession and expects high quality of doctors practicing medicine with respect for patients' integrity. Prevailing mechanisms by which various jurisdictions formally grant recognition of doctors as specialists are concentrated around three main strategies: 1) end-of-training examination 2) recognition on behalf of demonstrated clinical experience and years in training; and 3) in-training evaluation reports from the trainers.[7] Over time and jurisdictions these strategies have been more or less formalised and applied either singularly or in various combinations.[7] However, these strategies are now being questioned as to whether they are sufficient in

assuring trainees' competence according to society's needs, and new assessment strategies applied in the in-training context are requested.[1,8]

Internationally, professional bodies are increasingly responding to society's needs and various models specify the roles of future doctors and the expected learning outcomes. In essence these models are quite similar, see Table 1.[9-12]

Table 1. Prevailing models applying to a broad context-based concept of competence

	Harden, Crosby and Davis[a] 12 learning outcomes for a competent doctor	**JCHMT**[b] 12 generic aspects of the role of the consultant	**RCPSC**[c] 7 roles of the future specialist	**ACGME**[d] 6 general competencies for residents
Performance of the task	1. Clinical skills 2. Practical procedures 3. Patient investigation 4. Patient management 5. Health promotion 6. Communication 7. Information handling skills	1. Good clinical care 2. Communication skills 3. Information use and management 4. Cross specialty skills	1. Medical expert 2. Health promoter 3. Communicator	1. Patient care 2. Medical knowledge
Approach to the task	8. Understanding of social, basic and clinical sciences 9. Decision-making and clinical reasoning and judgement 10. Appropriate attitudes, ethical and legal responsibilities	5. Working with colleagues 6. Team working and leadership skills 7. Structure and principles of management 8. Teaching 9. Research 10. Clinical governance	4. Collaborator 5. Manager 6. Scholar	3. Interpersonal and communication skills 4. Practice-based learning and improvement
The professional in context	11. Role within the health service 12. Personal development	11. Maintaining trust 12. Maintaining good practice	7. Professional	5. Professionalism 6. System-based practice

[a] Harden, Crosby and Davis[8], [b] Joint Commission of Higher Medical Training[11], [c]Royal College of Physicians and Surgeons in Canada[14], [d]Accreditation Council of Graduate Medical Education[10].

Harden *et al.*[9] pictures a model of an outcome-based curriculum as three layers in a circle. The core represents the doctor's knowledge and skills, the middle layer the doctor's approach to the task, and the outer layer the doctor's role within a wider health service context and the continuing professional development.[9,13] In Canada a model of seven roles for the future specialist emerged,[10] in USA a model of 6 general competences,[11] and in UK a model of 12 generic aspects of the role as a consultant.[12] Postgraduate training programmes are now requested to define learning outcomes within these broad aspects of roles and elaborate assessment programmes accordingly.[11,12,14]

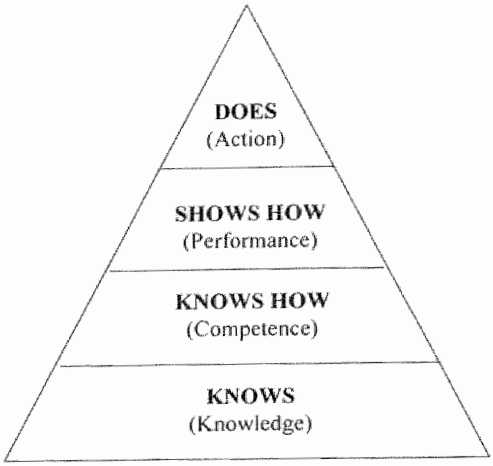
This introduction describes the rationale for in-training assessment (ITA) strategies in postgraduate education in light of the shifting concepts and strategies in medical education and the challenges of constructing ITA in a postgraduate work-based medical education context. In the first part the evolution of the concept of competence is described and why that calls for new assessment strategies. The second part addresses the problems of the prevailing strategies and why there is a need for effective in-training assessment strategies. The third part discusses the challenge of constructing in-training assessment programmes for a postgraduate work-based context. These challenges are seen from three different, but inter-related perspectives: the programme, the trainees, and the teachers.

In this thesis the term *in-training assessment* refers to various assessment instruments applied systematically in real clinical settings during training based on direct evidence such as observation or judgment of specific tasks completed by the trainee.[15] The term *in-training examination* refers to a knowledge test usually taken once or more during the training period and used formatively in contrast to the summative end-of training specialist exam.[16] The expression *in-training evaluation* is used here to refer to trainers' global rating of the trainees on various facets of clinical competence, which is usually applied by the end of defined period of time during the training.[16]

Shifting concepts and strategies

The concept of medical competence has evolved over the years from a narrow focus on only cognitive knowledge to also include skills and finally into the application of both in actual practice.[8,17] In undergraduate medical education the evolution of the concept of competence is reflected in the development of assessment strategies as pictured by an upward escalation in four layers of Miller's pyramid,[18] Figure 1.

Figure 1. Miller's pyramid of aspects of competence and assessment.



Framework for clinical assessment. Miller, 1990

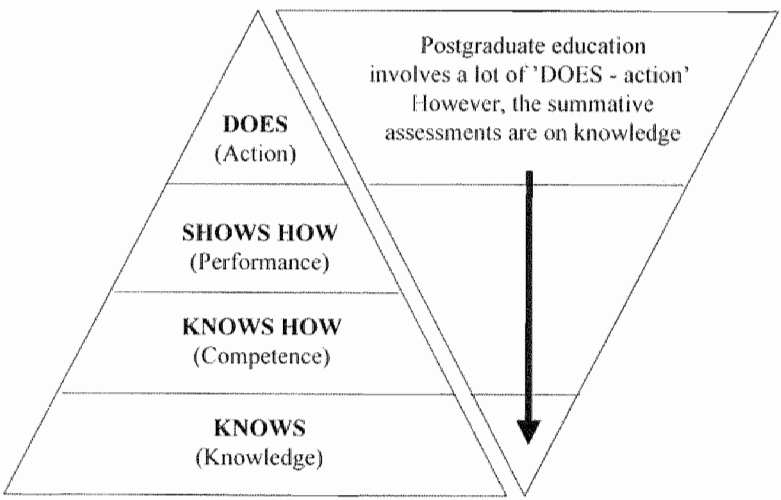
The bottom of the pyramid is 'Knows=knowledge', and for centuries that was the definition of professional competence as reflected in most examinations in the medical profession.[17] In undergraduate medical education the biggest innovation in assessment strategies in the last century was the introduction of the OSCE in 1975, which moved the monocular focus on competence as a cognitive function to also including rule-based skills.[19]

The second layer in the pyramid is 'Knows how=competence', and according to Miller this is 'the skill of acquiring information from a variety of sources, to analyse and interpret these data, and finally to translate such findings into a rational diagnostic or management plan'. The third layer is 'Shows how=performance in the artificial examination setting demonstrating the process through which conclusions are reached and not only the accuracy of diagnosis'. The distinction between competence and performance has been subject to some confusion and the terms are ambiguous.[17] Some authors define competence as what doctors know and can do under ideal circumstances, whereas performance refers to what is actually done under existing circumstances.[17,20] Others use the term performance about what is directly observable, whereas competence refers to something that is not directly observable, but inferred from performance.[21] In American literature the term competence is given a generic or holistic meaning and refers to an overall capacity, whereas the term competency refers to specific capabilities.[17] The holistic and generic aspect of competence is sometimes referred to as meta-competencies or general competencies.[11,13] The top layer in the pyramid 'Does=action' is

defined by Miller as 'the functioning independently in clinical practice'. Postgraduate education is work-based, which implies a lot of 'does=action'. In line with the future expectations of a wide variety of specialists' roles, Epstein and Hundert[22] define competence as "the habitual and judicious use of communication, knowledge, technical skills, clinical reasoning, emotions, values, and reflection in daily practice for the benefit of the individual and the community being served".

Irrespective of distinctions between competence and performance, the focus of postgraduate education is on the top layers of Miller's pyramid. Nevertheless, the prevailing formal assessment strategies in postgraduate education have been testing cognitive abilities applied as end-of training examinations, Figure 2. The recognition of a broad and context-based concept of competence implies that assessment of cognitive abilities is not enough and supplementation with assessment in real practice is recommended.[8,21,23,24] The mission of postgraduate education is to structure the development of the individual learner through increasingly complex contexts in order to progressively become an independent competent practitioner.[25-27] Accordingly, there is a need for in-training assessment strategies with emphasis on learning through assessment and on-the-job learning, assessment of progression towards defined outcomes, assessment of integrated abilities, assessment of different forms of medical knowledge, and assessment of self-directed learning skills.[23,28,29]

Figure 2. Miller's framework for assessment in undergraduate education as it applies to postgraduate education.



Miller's framework for assessment in undergraduate education

Prevailing strategies

The prevailing strategies to ensure the outcome of postgraduate education are end-of training examination, certification on behalf of years in training and clinical experience, and formal recognition of training programmes according to specified standards. The prevailing in training strategies used as indicators of clinical performance and monitoring progress are trainers' in-training evaluations and in-training examinations. This section addresses the validity of these strategies and some problems identified in previous research.

Prevailing assessment strategies

The prevailing assessment strategies in postgraduate education are testing cognitive abilities applied as end-of training examinations or in-training examinations, and using trainers' in-training evaluations as an indicator of performance in practice. The link between results on these strategies and more rigorous assessment of clinical performance is complex. In internal medicine results on multiple assessments of clinical encounters, the mini-CEX, demonstrate rather high correlations (from r=0.59 to r=0.79) with the corresponding scores on the ABIM evaluation form and with corresponding sub-sets of scorings on the in training examination (r=0.57 to r=0.47).[30] The ABIM scores in turn are only modestly or poorly correlated (from 0.31 to 0.00) with the results on in-training examinations and certification exam.[31] The poorest correlations are within aspects of procedural skills and interpersonal behaviour included in the evaluation form. In surgery the performance on examinations testing knowledge correlates poorly with assessments of technical skills and assessment of operative performance.[32,33] Several studies from a variety of specialties have demonstrated that trainers' evaluations are poorly related to objective tests of clinical skills and knowledge[34-41] and often these evaluations suffer from poor reliability.[42,43] This indicates that assessment strategies targeting performance in practice are necessary supplements to the cognitive tests.

In order to monitor the trainees' progress during training and to identify trainees at risk mock examinations are offered to the trainees as in-training examination. The value of monitoring trainees' progress by use of in-training examinations of medical knowledge might be different across specialties.[32,44] In internal medicine the most frequent reported difficulties with problem residents were insufficient knowledge (48%), poor clinical judgement (44%), and insufficient use of time (44%) (Yao).[45] In contrast a report from neurosurgery programmes indicate that problems were rarely attributed to cognitive or psychomotor deficits, rather dismissals were more often due to issues of professionalism, ethics, and

interpersonal skills and behaviours.[46] Attrition from postgraduate medical programmes remains a small but persistent issue and problem residents are a significant challenge to programme directors within most programmes.[45,47,48] ITA programmes with clear standards of performance within the wide aspects of competence and clear formulation of expected progress are suggested for the prevention, identification, management, and remediation of problem trainees.[45,49,50]

Certification examinations serve to ensure that doctors provide high quality of practice to the public. The link between certification and performance in practice varies considerably across studies.[24,51-53] A review of the literature from USA summarises the validity of certification in this respect.[24] In 16 out of 33 studies a positive association between certification status and clinical outcomes was demonstrated, 14 revealed no association, and three studies showed worse outcomes for certified doctors.[24] Several of the studies were problematic due to differences in case-mix among certified and non-certified doctors and various other contributors to the outcomes, which were not controlled for in the studies. In surgical specialties examinations focusing on knowledge cannot be expected to relate to technical ability.[53] In other specialties a straightforward relationship between certification and performance in practice is more logical for some aspects of competence. For example in family medicine performance on diagnosis and management sub-scores on the certifying examination was associated with higher rates of disease-specific relative to symptom-relief prescribing, and sub-scores on drug knowledge in the licensing examination were associated with a lower risk of contraindicated prescribing.[54] How recently doctors have graduated predicts performance in certain aspects of practice[55-57] indicating a need for continuous professional development after certification as a specialist.[58] Doctors must acquire the skills and attitudes towards continuously assessing and up-dating their own knowledge and skills according to the development of new medical technology as well as other aspects of competence.[58-60] Merely attending courses might not be sufficient in changing physicians' practice.[61] ITA programmes might contribute to fostering self-directed learning skills and attitudes.[23,28,29]

Clinical experience

In many countries and jurisdictions for example in Europe[62] there are no end-of training specialist exams. In these countries certification is usually granted in recognition of training in formal programmes or documentation of years spent in specified positions. Internationally there are various policies and guidelines concerning clinical experience and formal training.[11,12,14,63,64]

The designated number of years is quite arbitrary and varies considerably between jurisdictions as does the amount of working hours. An example from anaesthesiology is given in Table 2.

Table2. Minimum requirements of clinical experience in Denmark (DK), European Union (EU), United Kingdom (UK), and United States of America (USA)

	DK[a]	EU[c]	UK[a]	USA
Working hours per week	37½ hours	48-56 hours	72 hours	80 hours
Total length of training	**8½ years**		**8 years**	**4 years**
Internship	1½ years		1 year	1 year
Anaesthesiology, in total	**7 years[b]**	**5 years**	**7 years**	**3 years**
Basic anaesthesiology	12 months		24 months	12 months
Electives, other specialties	6 months		12 months	
Rotation requirements				
• Obstetric anaesthesia	2 months		3 months	1 month
• Paediatric anaesthesia	2 months		3 months	1 month
• Neuro-anaesthesia	2 months		3 months	1 month
• Thoracic-anaesthesia	2 months		3 months	1 month
• Anaesthesia for vascular surgery	1 month		3 months	
• Postoperative observation				2 weeks
• Intensive care medicine	12 months	6 months	3 months	2 months
• Pain treatment	3 months	3 months	3 months	1 month
• Emergency medicine, trauma	3 months	3 months		
• General anaesthesiology			6 months	
Additional experience in anaesthesiology	24 months[b]			

[a] DK and UK are both part of EU, which have an overall policy on maximum working hours. Each country in EU may have individual regulations regarding further limits on working hours. Each country has individual regulations for postgraduate medical training.
[b] Five years formal training is required to become certified as a specialist and additional two years experience in anaesthesiology is requested to become eligible to a consultant position.

The actual time spent in training within each year[65-67] is highly variable, and the proportion of time considered being of educational value in the residencies is between 11 and 76%.[68] Regulations that limit working hours of residents have globally been met with serious concerns about residents' clinical experience.[62,69-74] These aspects add to the reasons for

constructing effective postgraduate training programmes that ensure educational objectives are met and ITA might be one of the means to reach that end.[53] Long[75] demonstrated that structured ITA considerably increased the efficacy of postgraduate education with a 50% reduction in the time trainees' required to master procedures in neurosurgery.

Many specialties include a recommended caseload in their objectives. The recommended volume is not always in line with the learning curve for achieving competence.[76-79] Clinical experience measured by number of patients seen or number of procedures performed varies considerably among trainees.[80-82] In undergraduate education the amount of clinical experience is poorly related to results on either cognitive tests[83] or objective skills tests.[84,85] The benefit from clinical experience appears to be related to well-organised learning styles and the quality of teacher support.[52,84-89] In postgraduate education the importance of clinical experience and volume has repeatedly been demonstrated.[79,90-96] However, learning curves seems to be rather individual and for several procedures and performances it is unknown how much experience is sufficient for trainees to become competent. Formal ITA with clear performance standards rather than quantitative measures might be of help in constructing efficient training strategies and ensuring effective outcomes.[53]

Quality assurance of programmes

More or less formal recognition or accreditation of postgraduate training programmes serve the purpose of ensuring the quality in the learning environment at the training sites. This includes standards for the sponsor institutions and standards of general educational principles.[11,12,14,63,64] Historically these requirements were instituted when it became evident to the examiners that a substantial number of candidates coming up to final examinations had severe deficits in their clinical training.[64] Nowadays accreditation includes a considerable number of standards that programmes are expected to meet, which for example in internal medicine amounts to 301.[97] Among significant predictors of a programme's accreditation cycle length are total number of citations in the notification letter, percentage of trainees passing the certification exam, inadequate participation of trainees in teaching sessions and conferences, inadequate breadth in clinical experience, inadequate evaluation of trainees, and features of working conditions.[97]

A link between accreditation status and trainees' pass rates on end-of training examinations has been demonstrated.[97,98] Whether this is a causal relationship is a question as pass rates is one of the standards determining accreditation status. It has been demonstrated that doctors who have attended formal training programmes in internal medicine and in family medicine perform better on certifying exams and in prac-

tice.[40,56,57,96,99-103] However, pass rates on certification examination might not be the best indicator of the effect of formal training due to the competitive selection of trainees to programmes. Performance on prior examinations is a major contributor to the variance in performance on certification examinations.[40,100] When statistically controlling for the effect of trainees' performance on previous examinations, educational characteristics, like faculty/student ratios, length of training, performance during training, and well-supervised educational experience account for only 10-20% of the variance in scores on certification examination.[40,98,100]

Despite the emphasis accreditation bodies put on teaching sessions and conferences, the quality and effect of these activities on the learning outcome are questionable. Teaching sessions often fail to embrace standard educational principles such as needs assessment, program evaluation, and assessment of learning outcome.[104] Trainees' attendance does not correlate with scores on cognitive tests[105-107] unless interactive teaching formats are used or initiatives supporting trainees' self-study are included.[108-110]

Features of residents' working conditions such as service obligations or call-status do not significantly affect performance on in-training examinations.[111,112] Rather amount of study and learning style account for performance on cognitive tests.[109,110] However, work place characteristics such as high intensity of workload and poor staff support have repeatedly been shown to increase junior doctors' stress levels,[65,113-115] leading to a risk of getting into interpersonal conflicts and making serious medical errors.[65,116] Reforms in postgraduate education introducing formal requirements of structured programmes and teacher support may increase trainees' satisfaction with training and the likelihood of better induction, consultant supervision and feedback, and more hands on clinical experience.[117] Whether this also means that trainees become more competent has been questioned. In a study from England trainees' confidence in several procedures seemed to decline after the introduction of the Calman reform.[72]

Demonstrating clinical learning outcomes and quality of clinical performance while in training could be of help in comparing the quality of programmes. Several accreditation bodies now request programmes to apply rigorous ITA of learning outcomes and performance while in training with assessment of broad aspects of competence including interpersonal skills and behaviour.[11,12,14]

ITA programmes

The potential benefits of ITA in postgraduate education have been described and a wide variety of assessment instruments have been suggested for inclusion in comprehensive ITA programmes.[1,11,12,22] However, the literature on how ITA work in practice and what educational outcomes can actually be achieved is limited.[118] Any assessment programme can only cover a sub-set of the entire curriculum.[119,120] The challenge is to choose the important foci that cover the broad aspect of competence as it applies to each particular specialty and select appropriate assessment instruments that fit into a busy work-based postgraduate education context.[16,119,120] Secondly, it is a challenge to introduce the new concepts of competence and assessment in practice. These challenges are seen from three different, but inter-related perspectives: the programme, the trainees, and the teachers.

The programme

The utility of an assessment programme is based on the inter-related aspects of the programmes' validity, reliability, feasibility, acceptability, and educational impact.[119,121] The weight one should put on each of these aspects is often a trade-off and a compromise when keeping the overall purpose of the assessment programme in mind. In theory no assessment instrument can be valid unless it has sufficient reliability to be acceptable to the users and perceived as fair.[118,122] However, in the pursuit of reliable assessment instruments there is a risk of drawbacks concerning the validity, feasibility, acceptability, and effect on learning.[119,121] Having a reliable instrument that is not valid does not make sense. Strategic use of the integrated assessment programmes at the macro-level is recommended with an awareness of what educational impact it has as a total system.[119] Similarly studies of psychometric properties, feasibility, and acceptability must address the total system and not only the individual elements at the micro-level.[119] The psychometric properties of an assessment instrument can only be judged in terms of data collected from authentic application.[123] Thus an instrument is not in itself reliable or valid, it's the application of the instrument in a particular context in practice and the interpretation of the results obtained that has these properties.[123] Traditional strategies of psychometric judgment and standard setting apply to end-of training assessments with pooled data on groups of students, and usually expert assessors are involved.[124] However, these strategies do not readily apply to progressive assessment of individual trainees' mastery in clinical practice.[125] An ITA programme using mastery criteria for progress and the possibility of re-tests until mastery means that data would not include the variance traditionally used for psychometric analysis and

standard setting.[125] Therefore, the internal rational validation process becomes extremely important when constructing ITA programmes and introducing new and broader aspects of competence to be assessed.

In constructing an assessment programme definition of what is to be assessed is most important and the selection of appropriate methods comes second.[123] Sampling across many observations and using multiple assessors over time is recommended in order to ensure reliability.[42,119,111] Several instruments based on these principles have been suggested. However, very elaborate programmes bear the risk of bureaucratic overloading,[47,120,126] which reduce their feasibility and acceptability, especially if they are not educational and beneficial.[122,126,127] For example at least 30 report charts based on medical record audits are needed to achieve acceptable levels of generalisability.[128] Disappointingly, such report charts may be ineffective in improving residents' practice performance.[129] Another example is using peers and nurses to judge humanistic qualities, which requires sampling across 13 performance categories using 10-15 ratings from nurses or 11 ratings from peers.[130-132] Further, 20-40 patients' ratings of residents' communication skills on five to ten items have been shown to be sufficient for obtaining a reproducible, meaningful indicator.[133] Given the fact that all aspects of the future roles as a specialist should be covered in an assessment programme, instruments like these could add up to an immense workload on trainees, trainers and other health professionals.[126,134]

One problem with ITA is poor reliability when using clinicians as assessors.[135-137] Using checklists with task specific items might improve accuracy of assessment[135] and checklists might be seen as a help to the clinicians in making fair decisions on the trainees' performance. On the other hand checklists bear the risk of trivialising the content and global rating scales might be better in reproducibility as well as more efficient to develop and administer.[138,139] In standardised settings using expert assessors global rating scales has repeatedly been shown to have better psychometric properties than checklists.[138-142] Whether these findings also apply in a real life postgraduate setting has not been shown.[118] In postgraduate education, assessment instruments using general categories of competence have been shown to have sufficient reliability.[30,31,143] However, some of these instruments fail to demonstrate increasing levels of competence over the time of training and might not be suitable for monitoring trainees' progress or serve as a tool for learning and feedback.[30,31,118,143] Thus, in designing ITA programmes the challenge is to strike the balance between too much and too little if they are to have sufficient utility and applicability in a postgraduate work-based context.[16,119]

The learner

The learners in postgraduate education desire training in reflective practice, increasing in complexity over the years of training through ongoing supervision and feedback and on-the-job discussions with consultants.[144] Furthermore they want algorithms for medical decision-making, regular sign-offs on a competencies logbook, effective and strategic use of annual assessment, and development of more objective instruments for use in psychomotor skills training and assessment.[144] In other words comprehensive assessment programmes that foster learning and the development of competence in increasingly complex contexts are requested. Elaborating and introducing assessment strategies is a delicate balance between too much structure and too little. Trainees in postgraduate education are highly independent adult learners with a high motivation to learn and function as individual practitioners.[145] Too much rigor and structure might take away learners' initiatives and responsibility for own learning, and too little will make the learning haphazard and ineffective.[117,146] Assessment has profound effects on both learning and teaching through its content, format, feedback, and programming.[119] However, these effects are unpredictable and can be both positive and negative.[119,125]

Another problem of introducing formal assessment procedures is avoiding a blame and shame culture.[122,147] Resistance to formal assessment procedures can be expected especially in postgraduate contexts where examinations are not the tradition.[62] For example when formal assessment procedures were introduced for UK doctors, the attitude towards performance procedures changed in negative direction over time and experience with these.[147] The negative attitudes included that the procedures were perceived to foster more defensive practice, making doctors vulnerable, being unfair to some doctors, impairing medical morale and teamwork, and being principally window-dressing.[147] An ITA programme applied progressively through the training with adequate time to remediation in case of failures might overcome some of these barriers.[122]

The teachers

Apprenticeship learning and discussion with senior colleagues are major sources of learning in postgraduate education. However, the trainers might not always be up-to-date in knowledge and skills[59,148] and in case of disagreements over patient management decisions, discussions might be hampered due to the perception of knowledge deficits from both sides.[149,150] For better or for worse many of the medical and non-medical aspects of competence are learned through role modelling.[151,152] For example doctors' team working skills have been questioned and deficien-

cies in the collaboration with team members have been demonstrated in ward rounds, in the OR, and in the emergency room.[153-157] However, structuring collaboration between team members can be of benefit to patient care in addition to improving staff satisfaction, team spirit and communication.[155,156] Novices in a profession tend to mimic their role models and due to lack of experiential insight they might even misinterpret them.[158] Unfortunately this implicit learning is more memorable than the explicit curriculum.[156] Hence, explicit instruments for work-based teaching and assessing of both medical and non-medical aspects of competence might be of benefit to all parties.

Teachers' skills in clinical teaching can be of significance to trainees' learning out-come. In a study by Anderson *et al.*[159] students' performance on objective clinical tests were significantly correlated to their perception of the quality of instruction. This included making goals and responsibilities clear, being flexible to meet needs, having reasonable expectations, being accessible for discussion, giving supervision and feedback, asking challenging questions, and demonstrating concern for progress. ITA programmes specifying standards of performance and progress might be of help to clinical teachers in carrying out their instructor responsibilities in a busy working context.

Training the teachers for the role as assessors is recommended in order to minimise the assessor bias in the results.[118] ITA in practice is often opportunistic requiring almost any senior clinician to take on the role of being assessor. Elaborate ITA programmes might be of help to the programme managers in preparing the clinicians for the role as assessors.[118]

Summary of problem formulation

The recognition of a broad and context-based concept of competence implies that assessment of cognitive abilities is not enough and supplementation with assessment in real practice is recommended. In order to ensure structured development of the trainees into competent individual practitioners there is a need for using assessment as an on-going part of the learning process, and ITA strategies seem to be a means to meet that end. The special conditions for postgraduate education being based in a busy working context and not a teaching institution offer some unique challenges to those who are responsible for constructing assessment programmes. These challenges include compiling a programme that sufficiently covers the content and at the same time is feasible in practice without bureaucratic overloading of the users. Despite evidence that instruments using global scoring of competence are more efficient to develop and administer, there is a need for sufficiently elaborate tools to support the learning process and be of help to the trainers in providing feedback and monitoring progress. Finally, the programme should aim at

avoiding a blame and shame culture while serving its summative purpose of ensuring performance standards, quality of care and patient safety.

Research questions

This thesis focuses on the challenges of constructing and introducing an ITA programme, which is in accordance with the new concepts and methods introduced in medical education and applicable in a postgraduate work-based context. The overall research question of this dissertation is:

- How do theory on education and previous research outcomes inform the design and content of ITA in a work-based postgraduate medical education context and how does that apply to specific ITA programme in actual practice?

Six sub-questions were generated from the overall research question, which related to the aspects of design and content of an ITA programme, feasibility of ITA in a postgraduate context, and the educational impact of ITA. These research questions were applied to a specific context of 1st year training in anaesthesiology

Design and content of an ITA programme
In developing ITA that address new concepts and aspects of competence in postgraduate medical education, a careful internal rational validation process based on existing literature and theory becomes extremely important.[8,160] This process includes defining the content and selecting instruments that are supposedly applicable in a work-based context, and prospectively setting standards of performance and pass/fail criteria.[8,120,161] The first research question of the thesis addresses this aspect:

1. What is the internal rational validity of ITA in postgraduate education? How do theory on education and previous research outcomes inform the design and content of a specific ITA programme?

When new concepts and strategies are to be introduced, consensus on the content and the assessment protocol should be sought from content experts and those expected to be examiners. In ITA numerous clinicians will be the examiners. The second research question addresses this aspect of validation.

2. What is the content validity of ITA addressing broad aspects of competence? Do clinical experts agree on the composite content and the pre-defined standards for making decisions about the trainees?

Feasibility of an ITA programme
It is difficult to decide whether task specific checklists or global scoring forms using categories of competence are most appropriate in the postgraduate clinical context. The third research question addresses this aspect.

3. What is the feasibility of scoring forms in postgraduate education? Do clinicians prefer task specific checklists or global scoring forms with categories of competence when assessing trainees' performance? Is clinicians' agreement on pass/fail decisions better when using checklists compared to global scoring forms?

No matter how extensive the internal rational validation process and the prospective consensus seeking, barriers to implementing new concepts and imposing ITA in a busy work-based environment can be expected. Research question four addresses the feasibility of the programme in practice.

4. To what extent can ITA be implemented? How well is ITA accepted by the trainees? What are the positive and negative aspects?

Educational impact
The possible benefits and the educational impact of ITA are described in the literature. However, the literature on how ITA works in actual practice is sparse. Research question five addresses this aspect:

5. What is the educational impact of ITA in practice? How does it actually work in practice and how does that affect the training, teaching and learning?

Assessment drives learning and one of the stated aims of ITA is to increase trainees' clinical confidence. The last research question of this thesis address this aspect:

6. What is the effect of ITA in postgraduate education on trainees' clinical confidence?

The six research questions in this thesis relate to the construction and introduction of an ITA programme, and the first experience with the

programme. However, these studies are but a start and there are many other important questions related to ITA that have to be studied at a later stage. Those issues and future research questions are addressed in the concluding section of this thesis.

Structure of the thesis

The research questions in this thesis are addressed through six individual studies presented in each of six chapters. *Chapter one* deals with the internal rational validation process and how the theory on education and previous research outcomes informed the design and content of a specific ITA programme embracing broad aspects of competence. *Chapter two* presents the external validity of the programme, i.e. the clinical experts' opinion about the content of the programme and the pre-defined standards of performance and decisions that are to be made on behalf of the programme. *Chapter three* is a study of the feasibility of task specific checklists vs. global rating forms addressing general aspects of competence for the purpose of assessing trainees' clinical performance in postgraduate education. *Chapter four* is a study of the feasibility of the programme in practice, and preliminary data on the trainees' acceptance of ITA. *Chapter five* is a qualitative study of how the programme works in actual practice and the effect on training, teaching and learning. *Chapter six* is a study of the effect of ITA on trainees' clinical confidence. Since this dissertation is based on journal articles some information pertaining to the ITA programme itself may be repetitive.

References

1. Murray E, Gruppen, L, Catton P, Hays R, Woolliscroft JO. The accountability of clinical education: its definition and assessment. Med Educ 2000; 34: 871-9.
2. Kilminster SM, Jolly BC. Effective supervision in clinical practice settings: a literature review. Med Educ 2000; 34: 827-40.
3. McCue. Addressing the service versus education conflicts created by residency training in ambulatory care settings. J Ambul Manage 1995; 18: 84-90.
4. Aitken RJ, Thompson MR, Smith JAE, Radcliffe AG, Stamatakis JD, Steele RJC. Training in large bowel cancer surgery: observations from three prospective regional United Kingdom audits. BMJ 1999; 318: 702-3.
5. Baldwin PJ, Newton RW, Buckley G, Roberts MA, Dodd M. Senior house officers in medicine: postal survey of training and work experience. BMJ 1997; 314: 740-3.
6. Freeman ML. Training and competence in gastrointestinal endoscopy. Rev Gastroenterol Disord 2002; 1: 73-86.
7. Dauphinee WD. Licensure and certification. In: Norman, GR, Van der Vleuten, CPM, Newble, DI, eds. International Handbook of Research in Medical Education, Dordrecht: Kluwer Academic Publishers. Printed in Great Britain. 2002: 835-79.

8. Turnbull J, Van Barneveld C. Assessment of clinical performance: In-training evaluation. In: Norman, GR, Van der Vleuten, CPM, Newble, DI, eds. International Handbook of Research in Medical Education, Dordrecht: Kluwer Academic Publishers. Printed in Great Britain. 2002: 793-810.
9. Harden RM, Crosby JR, Davis MH. AMEE Guide No. 14: Outcome-based education: Part 1 - An introduction to outcome-based education: a model for the specification of learning outcomes. Med Teach 1999; 21: 7-14.
10. CanMEDS 2000: Extract from the CanMEDS 2000 Project Societal Needs Working Group Report. Med Teach 2000; 22: 549-54.
11. ACGME Outcome project. Accreditation Council for Graduate Medical Education Web site: http://www.acgme.org. 2000. Accessed April 2003.
12. JCHMT: Higher medical training. Generic curriculum. Joint Committee for Higher Medical Training Web site: http://www.jchmt.org.uk. 2003. Accessed April 2003.
13. Harden RM, Crosby JR, Davis MH, Friedman M. AMEE Guide No. 14: Outcome-based education: Part 5 - From competency to meta-competency: a model for the specification of learning outcomes. Med Teach 1999; 21: 546-52.
14. RCPSC General standards of accreditation. Royal College of Physicians and Surgeons in Canada Web site: http://www.rcpsc.medical.org. 2002. Accessed April 2003.
15. Feletti G, Cameron D, Dawson-Saunders B, Des Groseilliers J, Dooley B, Farmer E, McAvoy P. In-training assessment. In: Newble D, Jolly B, Wakeford R, eds. The certification and recertification of doctors: Issues in the assessment of clinical competence. Cambridge University Press. 1994: 151-66.
16. Holmboe ES, Hawkins RE. Methods for evaluating the clinical competence of residents in internal medicine: a review. Ann Intern Med 1998; 129: 42-8.
17. Eraut M. Concepts of competence and their implications. In: Eraut M, ed. Developing professional knowledge and competence. The Falmer press. London. 1999: 163-81.
18. Miller GE. The assessment of clinical skills/competence/performance. Acad Med 1990; 65(Suppl): S63-7.
19. Harden R, Gleeson F. Assessment of clinical competence using an objective structured clinical examination (OSCE). Med Educ 1979; 13: 41-54.
20. Rethans J-J, Norcini JJ, Barón-Maldonado M, Blackmore D, Jolly BC, LaDuca T, Lew S, Page GG, Southgate LH. The relationship between competence and performance: implications for assessing practice performance. Med Educ 2002; 36: 901-9.
21. Haeger P, Gonczi A. What is competence? Med Teach 1996; 18: 15-8.
22. Epstein RM, Hundert EM. Defining and assessing professional competence. JAMA 2002; 287: 226-35.
23. Friedman Ben-David M. The role of assessment in expanding professional horizons. Med Teach 2000; 22: 472-6.
24. Sharp LK, Bashook PG, Lipsky MS, Horowitz SD, Miller SH. Specialty board certification and clinical outcomes: The missing link. Acad Med 2002; 77: 534-42.
25. Leach DC. Competence is a habit. JAMA 2002; 287: 243-4.
26. Eraut M. Theories of professional expertise. In: Eraut M, ed. Developing professional knowledge and competence. The Falmer press. London. 1999: 123-57.
27. Dreyfus HL, Dreyfus SE. Mind over machine: The power of human intuition and expertise in the era of the computer. Oxford, Basil Blackwell. 1986.
28. Dauphinee WD. Assessing clinical performance. Where do we stand and what might we expect? JAMA 1995; 274: 741-3.
29. Shepard LA. The role of assessment in a learning culture. Educ Research 2000; 29: 4-14.
30. Durning SJ, Lannie JC, Markert RJ, Pangaro LN. Assessing the reliability and validity of the mini-clinical evaluation exercise for internal medicine residency training. Acad Med 2002; 77: 900-4.
31. Durning SJ, Cation LJ, Jackson JL. The reliability and validity of the American Board of Internal Medicine monthly evaluation form. Acad Med 2003; 78: 1175-82.

32. Scott DJ, Valentine RJ, Bergen PC, Rege RV, Laycock R, Tesfay ST, Jones DB. Evaluating surgical competency with the American Board of Surgery in-training examination, skill testing, and intraoperative assessment. Surgery 2000; 128: 613-22.
33. Sloan DA, Donnelly MB, Schwartz RW, Strodel WE. The objective structured clinical examination. The new gold standard for evaluating postgraduate clinical performance. Ann Surg 1995; 222: 735-42.
34. Sloan DA, Donnelly MB, Johnson SB, Schwartz RW, Strodel WE. Use of an objective structured clinical examination (OSCE) to measure improvement in clinical competence during the surgical internship. Surgery 1993; 114: 343-51.
35. Fox RA, Clark CLI, Scotland AD, Dacre JE. A study of pre-registration house officers' clinical skills. Med Educ 2000; 34: 1007-12.
36. Dupras DM, Li JTC. Use of an objective structured clinical examination to determine clinical competence. Acad Med 1995; 70: 1029-34.
37. Kahn MJ, Merrill WW, Anderson DS, Szerlip HM. Residency program director evaluations do not correlate with performance on a required 4th-year objective structured clinical examination. Teach Learn Med 2001; 13: 9-12.
38. Joorabchi B. Objective structured clinical examination in a pediatric residency program. AJDC 1991; 145: 757-62.
39. Hawkins RE, Sumption KF, Gaglione MM, Holmboe ES. The in-training examination in internal medicine: resident perceptions lack of correlation between resident scores and faculty predictions of resident performance. Am J Med 1999; 106: 206-10.
40. Norcini JJ, Grosso LJ, Shea JA, Webster GD. The relationship between features of residency training and ABIM certifying examination performance. J Gen Intern Med 1987; 2: 330-6.
41. Wise S, Stagg PL, Szues R, Gay S, Mauger D, Hartman D. Assessment of resident knowledge: subjective assessment versus performance on the ACR in-training examination. Acad Radiol 1999; 6: 66-71.
42. Kreiter CD, Ferguson K, Lee W-C, Brennan RL, Densen P. A generalizability study of a new standardized rating form used to evaluate students' clinical clerkship performances. Acad Med 1998; 73: 1294-8.
43. Gray JD. Global rating scales in residency education. Acad Med 1996; 71(Suppl): S55-63.
44. Garibaldi RA, Subhiyah R, Moore ME, Waxman H. The in-training examination in internal medicine: an analysis of residents' performance over time. Ann Intern Med 2002; 137: 505-10.
45. Yao DC, Wright SM. National survey of internal medicine residency programme directors regarding problem residents. JAMA 2000; 284: 1099-1104.
46. Cusimano MD, Yonke AM, Tucker WS. An analysis of attrition from Canadian neurosurgery residency programs. Acad Med 1999; 74: 925-31.
47. Beasley BW, Kern DE, Kolodner K. Job turnover and its correlates among residency program directors in internal medicine: a three-year cohort study. Acad Med 2001; 76: 1127-35.
48. Van Zanten M, Boulet JR, McKinley D, Whelan GP. Attrition rates of residents in postgraduate training programs. Teach Learn Med 2002; 14: 175-7.
49. Short JP. The importance of strong evaluation standards and procedures in training residents. Acad Med 1993; 68: 522-5.
50. Bergen PC, Littlefield JH, O'Keefe GE, Rege RV, Anthony TA, Kim LT, Turnage RH. Identification of high-risk residents. J Surg Res 2000; 92: 239-44.
51. Kramer AW, Jansen JJ, Zuithoff P, Dusman H, Tan LH, Grol RP, van der Vleuten CP. Predictive validity of a written knowledge test of skills for an OSCE in postgraduate training for general practice. Med Educ 2002; 36: 812-9.
52. Hutchinson L, Aitken P, Hayes T. Are medical postgraduate certification processes valid? A systematic review of the published evidence. Med Educ 2002; 36: 73-91.

53. Moorthy K, Munz Y, Sarker SK, Darzi A. Objective assessment of technical skills in surgery. BMJ 2003; 327: 1032-7.
54. Tamblyn R, Abrahamowicz M, Dauphinee WD, Hanley JA, Norcini J, Girard N, Grand'Maison P, Brailowsky C. Association between licensure examination scores and practice in primary care. JAMA 2002; 288: 3019-26.
55. Spike NS, Veitch PC. Analysis of the RACGP Fellowship examination. Aust Family Physician 1990; 19: 767-75.
56. Levy AR, Tamblyn RM, Mcleod PJ, Fitchett D, Abrahamowicz M. The effect of physicians' training on prescribing beta-blockers for secondary prevention of myocardial infarction in the elderly. Ann Epidemiol 2002; 12: 86-9.
57. Norcini JJ, Kimball HR, Lipner RS. Certification and specialization: Do they matter in the outcome of acute myocardial infarction? Acad Med 2000; 75: 1193-8.
58. Rethans J-J, Norcini JJ, Barón-Maldonado M, Blackmore D, Jolly BC, LaDuca T, Lew S, Page GG, Southgate LH. The relationship between competence and performance: implications for assessing practice performance. Med Educ 2002; 36: 901-9.
59. Vernooij-Dassen MJ, Ram PM, Brenninkmeijer WJ, Franssen LJ, Bottema BJ, van der Vleuten CP, Grol RP. Quality assessment in general practice trainers. Med Educ 2000; 34: 1001-6.
60. Guest CB, Regehr G, Tiberius RG. The life long challenge of expertise. Med Educ 2001; 35: 78-81.
61. Davis DA, Thompson MA, Oxman AD, Haynes RB. Changing physician performance. A systematic review of the effects of continuing medical education strategies. JAMA 1995; 274: 700-5.
62. Karle H, Nystrup J. Comprehensive evaluation of specialist training: an alternative to board examinations in Europe? Med Educ 1995; 29: 308-16.
63. Borman E, O'Grady P. Postgraduate training policy paper prepared by the subcommittee on postgraduate training: Permanent working group of European junior hospital doctors. Med Educ 1997; 31: 3-8.
64. Cassie, JM, Armbruster JS, Bowmer MI, Leach DC. Accreditation of postgraduate medical education in the United State and Canada: a comparison of two systems. Med Educ 1999; 33: 493-8.
65. Baldwin DC, Daugherty SR, Tsai R, Scotti MJ. A national survey of residents' self-reported work hours: Thinking beyond specialty. Acad Med 2003; 78: 1154-63.
66. Hoekelman RA, Parker RM. Variations in length of training among residents in paediatrics and internal medicine training programs. Acad Med 1990; 65: 257-66.
67. Pickersgill T. The European working time directive for doctors in training. BMJ 2001; 323: 1266.
68. Boex JR, Leahy PJ. Understanding residents' work: Moving beyond counting hours to assessing educational value. Acad Med 2003; 78: 939-44.
69. Directive 2000/34/EC of the European Parliament and Council. Official Journal of the European Community 2000; L195: 41-45.
70. Johnson T. Limitations on residents' working hours at New York teaching hospitals: a status report. Acad Med 2003; 78: 3-8.
71. MacLellan A. Residents' duty hours in the province of Quebec, Canada. Acad Med 2003; 78: 11-3.
72. Gillard JH, Dent TH, Smyth-Pigott PJ, Eaton J. Recent changes in the workload and clinical experience of pre-registration house officers: experiences over four years in south-west England. Med Educ 2000; 34: 371-3.
73. Kapur N, House A. Working patterns and the quality of training of medical house officers: evaluating the effect of the 'new deal'. Med Educ 1998; 32: 432-8.
74. Urowitz M, Crescenzi A, Muharuma L. Residents' duty hours in the province of Ontario, Canada. Acad Med 2003; 78: 9-10.
75. Long DM. Competency-based residency training: the next advance in graduate medical education. Acad Med 2000; 75: 1178-83.

76. Jowell PS, Baillie J, Branch S, Affronti J, Browning CL, Bute BP. Quantitative assessment of procedural skills. A prospective study of training in endoscopic retrograde cholangiopancreatography. Ann Intern Med 1996; 125: 983-9.
77. Konrad C, Scüpfer G, Wietlishbach M, Gerber H. Learning manual skills in anaesthesiology: Is there a recommended number of cases for anesthetic procedures? Anesth Analg 1998; 86: 635-9.
78. Kopacz DJ, Neal JM, Pollock J. The regional anaesthesia 'learning curve'. What is the minimum number of epidural and spinal blocks to reach consistency? Reg Anesth 1996;21:182-90.
79. Marshall JB. Technical proficiency of trainees performing colonoscopy: a learning curve. Gastraointest Endosc 1995; 42: 287-91.
80. Sandermann J, Panduro JL. The Danish specialist training in vascular surgery. Eur J Vasc Endocasc Surg 2002; 23: 353-7.
81. Nixon MC. The anaesthetic logbook – a survey. Anaesthesia 2000; 55: 1076-80.
82. Mangione S, Peitzman SJ. Revisiting physical diagnosis during the medical residency: It is time for a logbook - and more. Acad Med 1999; 74: 467-9.
83. Greenberg LW, Getson P. Clinical experiences of medical students in pediatric clerkship and performance on the NBME subject test: are they related? Med Teach 1999; 21: 420-3.
84. Martin IG, Stark P, Jolly B. Benefiting from clinical experience: the influence of learning style and clinical experience on performance in an undergraduate objective structured clinical examination. Med Educ 2000; 34: 530-4.
85. Chatenay M, Maguire T, Skakun E, Chang G, Cook D, Warnock GL. Does volume of clinical experience affect performance of clinical clerks on surgery exit examinations? Am J Surg 1996; 172: 366-72.
86. Godellas CV, Huang R. Factors affecting performance on the American Board of Surgery in-training examination. Am J Surg 2001; 181: 294-6.
87. Marteau TM, Wynne G, Kaye W, Evans TR. Resuscitation: experience without feedback increases confidence but not skill. BMJ 1990; 300: 849-50.
88. McManus IC, Richards P, Winder BC, Sproston KA. Clinical experience, performance in final examinations, and learning style in medical students: prospective study. BMJ 1998; 316: 345-50.
89. Paice E, Moss F, West G, Grant J. Association of use of a logbook and experience as a preregistration house officer: interview survey. BMJ 1997; 314: 213-5.
90. Martling LA, Holm T, Rutqvist L-E, Moran BJ, Heald RJ, Cedermark B, for the Stockholm Colorectal Cancer Study Group and the Basingstoke Bowel Cancer Research Project. Lancet 2000;356:93-6.
91. Martling A, Cedermark B, Johansson H, Rutqvist LE, Holm T. The surgeon as a prognostic factor after the introduction of total mesorectal excision in the treatment of rectal cancer. Br J Surg 2002; 89: 1008-13.
92. Novick RJ, Stitt LW. The learning curve of an academic cardiac surgeon: Use of the CUSUM method. J Card Surg 1999; 14: 312-20.
93. Norcini JJ, Lipner RA. Recertification: is there a link between take-home and proctored examinations? Acad Med 1999; 74(Suppl): S27-30.
94. Norcini JJ, Lipner RS. The relationship between the nature of practice and performance on a cognitive examination. Acad Med 2000; 75(Suppl): S68-70.
95. Shea JA, Norcini JJ, Baranowski RA, Langdon LO, Popp RL. A comparison of video and print formats in the assessment of skill in interpreting cardiovascular motion studies. Eval Health Prof 1992; 15: 325-40.
96. Steel K, Norcini J, Brummel-Smith K, Erwin D, Markson L. The first certifying examination in geriatric medicine. J Am Geriatr Soc 1989; 37: 1188-91.
97. Beasley BW, Scrase DR, Schultz HJ. Determining the predictors of internal medicine residency accreditation. What they do (not what they say). Acad Med 2002; 77: 238-46.

98. Wolfshal SD, Beasley BW, Kopelman R, Stickley W, Gabryel T, Kahn MJ, Membership survey and scientific data committee, the association of program directors of internal medicine. Benchmarks of support in internal medicine residency training programs. Acad Med 2002; 77: 50-6.
99. Norcini JJ, Shea JA, Langdon LO, Hudson LD. First American Board of Internal Medicine critical care examination: process and results. Crit Care Med 1989; 17: 695-8.
100. Norcini JJ. Indicators of educational effectiveness of subspecialty training programs in internal medicine. Acad Med 1995; 70: 512-6.
101. Norcini JJ, Lipner RS, Kimball HR. Certifying examination performance and patient outcomes following acute myocardial infarction. Med Educ 2002; 36: 853-9.
102. Borgiel AE, Williams JI, Bass MJ, Dunn EV, Evensen MK, Lamont CT, MacDonald PJ McCoy JM, Spasoff RA. Quality of care in family practice: does residency training make a difference? CMAJ 1989; 140: 1035-43.
103. Maheux B, Beaudoin C, Jacques A, Lambert J, Levesque A. Effect of residency training in family medicine v. internship training on professional attitudes and practice patterns. CMAJ 1992;146:901-7.
104. Hebert RS, Wright SM. Re-examining the value of medical ground rounds. Acad Med 2003; 78: 1248-52.
105. FitzGerald JD, Wenger NS. Didactic teaching conference for IM residents: who attends, and is attendance related to medical certifying examination scores? Acad Med 2003; 78: 84-9.
106. Picciano A, Winter R, Ballan D, Birnberg B, Jacks M, Laing E. Resident acquisition of knowledge during a noontime conference series. Fam Med 2003; 35: 418-22.
107. Pollak R, Baker RJ. The acquisition of factual knowledge and the role of the didactic conference in a surgical residency program. Am Surg 1988; 54: 531-4.
108. Bull DA, Stringham JC, Karwande SV, Neumayer LA. Effect of a resident self-study and presentation program on performance on the thoracic surgery in-training examination. Am J Surg 2001; 181: 142-4.
109. Hirvela ER, Becker DR. Impact of programmed reading on ABSITE performance. American Board of Surgery In-Training Examination. Am J Surg 1991; 162: 487-90.
110. Itani KM, Miller CC, Church HM, McCollum CH. Impact of a problem-based learning conference on surgery residents' in training exam (ABSITE) scores. American Board of Surgery In Training Exam. J Surg Res 1997; 70: 66-8.
111. Riggs JW, Johnson C, O'Niel P, Berens P. Are residents' work schedules related to their in-training examination scores? Obstet Gynecol 1996; 88: 891-4.
112. Stone MD, Doyle J, Bosch RJ, Bothe A, Steele G. Effect of resident call status on ABSITE performance. Surgery 2000; 128: 465-71.
113. Williams S, Dale J, Glucksman E, Wellesley A. Senior house officers' work related stressors, psychological distress, and confidence in performing clinical tasks in accident and emergency: a questionnaire study. BMJ 1997; 314: 713-8.
114. Tyssen R, Vaglum P, Grønvold NT, Ekeberg Ø. The impact of job stress and working conditions on mental health problems among junior house officers. A nationwide Norwegian prospective cohort study. Med Educ 2000; 34: 374-84.
115. Paice E, Rutter H, Wetherell M, Winder B, McManus IC. Stressful incidents, stress and coping strategies in the pre-registration house officer year. Med Educ 2002; 36: 56-65.
116. Gaba DM, Howard SK. Fatigue among clinicians and the safety of patients. N Engl J Med 2002; 347: 1249-55.
117. Paice E, Aitken M, Cowan G, Heard S. Trainee satisfaction before and after the Calman reforms of specialist training: questionnaire survey. BMJ 2000; 320: 832-6.
118. Spike NS, Veitch PC. Analysis of the RACGP Fellowship examination. Aust Family Physician 1990; 19: 767-75.
119. Van der Vleuten CPM. The assessment of professional competence: developments, research and practical implications. Adv Health Sci Educ 1996: 1: 41-67.

120. Schuwirth LWT, Southgate L, Page GG, Paget NS, Lescop JMJ, Lew SR, Wade WB, Borón-Maldonado M. When enough is enough: a conceptual basis for fair and defensible practice performance assessment. Med Educ 2002; 36: 925-30.
121. Van der Vleuten CPM, Scherpbier AJJA, Dolmans DHJM, Schuwirth LWT, Verwijnen GM, Wolfhagen HAP. Clerkship assessment assessed. Med Teach 2000; 22: 592-600.
122. Finucane PM, Barron SR, Davies HA, Hadfield-Jones RS, Kaigas TM. Towards an acceptance of performance assessment. Med Educ 2002; 36: 959-64.
123. Dauphinee WD, Blackmore DE. Assessing the assessors' assessment. Med Educ 2001; 35: 317-8.
124. Norcini J, Guille R. Combining tests and setting standards. In: Norman GR, Van der Vleuten CPM, Newble DI, eds. International Handbook of Research in Medical Education, Dordrecht: Kluwer Academic Publishers. Printed in Great Britain. 2002: 811-34.
125. Petrusa ER. Clinical performance assessments. In: Norman GR, Van der Vleuten CPM, Newble DI, eds. International Handbook of Research in Medical Education. Kluwer Academic Publishers, 2002: 673-709.
126. Leung W-C. Competency based medical training: review. BMJ 2002; 325: 693-6.
127. Hauer KE, Teherani A, Wiese J, Fenton CL. A strategy to standardize the learning of core clerkship objectives. Adv Health Sci Educ 2003; 8: 213-21.
128. Norcini JJ Jr. Standards and reliability in evaluation: When rules of thumb don't apply. Acad Med 1999; 74: 1088-90.
129. Kogan JR, Reynolds EE, Shea JA. Effectiveness of report cards based on chart audits of residents' adherence to practice guidelines on practice performance: A randomized controlled trial. Teach Learn Med 2003; 15: 25-30.
130. Ramsey PG, Wenrich MD, Carline JD, Inui TS, Larson EB, LoGerfo JP. Use of peer ratings to evaluate physician performance. JAMA 1993; 269: 1655-60.
131. Ramsey PG, Carline JD, Blank LL, Wenrich MD. Feasibility of hospital-based use of peer ratings to evaluate the performances of practicing physicians. Acad Med 1996; 71: 364-70.
132. Wenrich MD, Carline JD, Giles LM, Ramsey PG. Ratings of the performances of practicing internists by hospital-based registered nurses. Acad Med 1993; 68; 680-7.
133. Swanson, DB, Webster GD, Norcini JJ. Precision of patient ratings of residents' humanistic qualities: how many items and patients are enough? In: Bender W, Hiemstra RJ, Scherpbier AJJA, Zwierstra RP, (Eds.) Teaching and Assessing Clinical Competence. Groningen: BoekWerk Publications, 1990: 424-31.
134. Dunnington GL, Williams RG. Addressing the new competencies for residents' surgical training. Acad Med 2003; 78: 14-21.
135. Noel GL, Herbers JE, Caplow MP, Cooper GS, Pangaro JN, Harvey J. How well do internal medicine faculty members evaluate the clinical skills of residents? Ann Int Med 1992; 117: 757-65.
136. Herbers JE, Noel GL, Cooper GS, Harvey J, Pangaro LN, Weaver MJ. How accurate are faculty evaluations of clinical competence? J Gen Intern Med 1989; 4: 202-8.
137. Kroboth FJ, Hanusa BH, Parker S, Coulehan JL, Kapoor WN, Brown FH, Karpf M, Levey GS. The inter-rater reliability and internal consistency of a clinical evaluation exercise. J Gen Int Med 1992; 7: 174-9.
138. Govaerts MJB, Van der Vleuten CPM, Schuwirth LWT. Optimising the reproducibility of a performance-based assessment test in midwifery education. Adv Health Sci Educ 2002; 7: 133-45.
139. Regehr G, MacRae H, Reznick RK, Szalay D. Comparing the psychometric properties of checklists and global rating scales for assessing performance on an OSCE-format examination. Acad Med 1998; 73: 993-7.
140. Hodges B, Regehr G, Hanson M, McNaughton N. Validation of an objective structured clinical examination in psychiatry. Acad Med 1998; 73: 910-2.
141. Hodges B, Regehr G, McNaughton N, Tiberius R, Hanson M. OSCE checklists do not capture increasing levels of expertise. Acad Med 1999; 74: 1129-34.

142. Daelmans HEM, Scherpbier AJJA, Van der Vleuten CPM, Konker AJM. Reliability of clinical oral examinations re-examined. Med Teach 2001; 23: 422-4.
143. Norcini JJ, Blank LL, Arnold GK, Kimball HR. Examiner differences in the Mini-Cex. Adv Health Sci Educ 1997; 2: 27-33.
144. Khera N, Stroobant J, Primhak RA, Gupta R, Davies H. Training the ideal hospital doctor: the specialist registrars' perspective. Med Educ 2001; 35: 957-66.
145. Greaves JD, Grant J. Watching anaesthetists work: using the professional judgement of consultants to assess the developing clinical competence of trainees. Br J Anaesth 2000; 84: 525-33.
146. Calman KC, Temple JG, Naysmith R, Cairncross RG, Bennett SJ. Reforming higher specialist training in the United Kingdom - a step along the continuum of medical education. Med Educ 1999; 33: 28-33.
147. McManus IC, Winder BC, Gordon D. UK doctors' attitudes to the General Medical Council's performance procedures, 1997-99. Med Educ 2001; 35(Suppl.1): 60-9.
148. Iirola T, Lund VE, Katila AJ, Mattila-Vuori A, Pälve H. Teaching hospital physicians' skills and knowledge of resuscitation algorithms are deficient. Acta Anaesthesiol Scand 2002; 46: 1150-4.
149. Coats RD, Burd RS. Intraoperative communication of residents with faculty: perception versus reality. J Surg Res 2002; 104: 40-5.
150. Shreves JG, Moss AH. Residents' ethical disagreements with attending physicians: an unrecognized problem. Acad Med 1996; 71: 1103-5.
151. Paice E, Heard S, Moss F. How important are role models in making good doctors? BMJ 2002; 325: 707-10.
152. Prideaux D, Alexander H, Bower A, Dacre J, Haist S, Jolly B, Norcini J, Roberts T, Rothman A, Rowe R, Tallett S. Clinical teaching: maintaining an educational role for doctors in the new health care environment. Med Educ 2000; 34: 820-6.
153. Lindgard L, Reznick R, DeVito I, Espin S. Forming professional identities on the health care team: discursive constructions of the 'other' in the operating room. Med Educ 2002; 36: 728-34.
154. Elliot DL, Hickam DH. Attending ward rounds on in-patient units: differences between medical and non-medical services. Med Educ 1993; 27: 503-8.
155. Zwarenstein M, Bryant W. Interventions to promote collaboration between nurses and doctors (Cochrane Review). In: The Cochrane Library, Issue 4. Oxford: Update Software; 2001.
156. Risser D, Simon R, Rice M, Salisbury M. The potential for improved teamwork to reduce medical errors in the emergency department. Ann Emergency Med 1999; 34: 373-83.
157. Busby A, Gilchrist B. The role of the nurse in the medical ward round. J Adv Nurs 1992; 17: 339-46.
158. Lindgard L, Reznick R, Espin S, Regehr G, DeVito I. Team communications in the operating room: talk patterns, sites of tension, and implications for novices. Acad Med 2002; 77: 37-42.
159. Anderson DC, Harris IB, Allen S, Satran L, Bland CJ, Davis-Feickert JA, Poland GA, Miller WJ. Comparing students' feedback about clinical instruction with their performances. Acad Med 1991; 66: 29-34.
160. Fowell SL, Bligh JG. Recent developments in assessing medical students. Postgrad Med J 1998; 74: 18-24.
161. Hays RB, Davies HA, Beard JD, Caldon LJM, Farmer EA, Finucane PM, McCrorie P, Newble DI, Schuwirth LWT, Sibbald GR. Selecting performance assessment methods for experienced physicians. Med Educ 2002; 36: 910-7.

CHAPTER 1

Designing an in-training assessment (ITA) programme

Published in Medical Teacher 2003; 25: 54-62. *http://tandf.co.uk/journals*

Embracing the new paradigm of assessment in residency training. An assessment programme for first year residency training in anaesthesiology

C. Ringsted, D. Østergaard, and A. Scherpbier

Abstract

Assessment of clinical competence is facing a paradigm shift in more than one sense. The shift relates to test content, which increasingly covers a broader spectrum of competencies than mere medical expertise and to test methods, with an increasing focus on testing performance in realistic settings. Also there is a shift in the concept of assessment in that instruction and assessment are no longer seen as being separate in time and purpose, but as integral parts of the learning process. The nature of the new paradigm for assessment is well described, but the challenge to programme directors is to specify the evaluation situations and develop appropriate methods. This paper describes the intrinsic rational validation process in outlining an assessment programme for first year anaesthesiology residency training according to the new paradigm. The applicability to other residency programmes and higher-level training in anaesthesiology is discussed.

Introduction

Postgraduate assessment of clinical competence is facing a paradigm shift due to changes in the roles of physicians and of assessment. First, the content of assessment covers an increasingly broader spectrum of competencies than mere medical expertise. The Report of the CanMEDs 2000 project, for example, defines seven roles and competencies for future specialists: medical expert, scholar, communicator, health advocate, manager, collaborator and professional.[1] Similarly, The Accreditation Council for Graduate Medical Education in the United States has defined the following six general competencies: patient care, medical knowledge, practice-based learning and improvement, interpersonal and communication skills, professionalism, and system-based practice.[2] The challenges are to define and specify the evaluation situations and train the evaluators.[3] Second, the paradigm shift manifests itself in the assessment methods, which are targeting higher levels of Miller's pyramid.[4] From a previous focus on the lower levels of the pyramid, i.e. 'knows' and 'knows how', the emphasis is shifting to performance assessment where a candidate must 'show how' and/or be observed 'doing' the tasks reflecting the competencies at stake.[5] Considerable experience and high psychometric quality have been achieved in assessment at the level of 'shows how', with standardised formats like the observed structured clinical examination (OSCE) and simulated patients.[6, 7] At the 'does' level the prevailing method in residency assessment has been in-training observation with scoring on rating scales. There have been serious problems identified in

the reliability of this approach.[8, 9] Finally, there is a marked shift in the concept of assessment. Whereas instruction and assessment used to be separated in time and purpose, the current trend is towards assessment as an integral part of the learning process.[10] If we want to ensure that assessment does indeed drive learning and constitutes a learning experience, a variety of assessment methods is needed along with continuous documentation of the learning of the resident. The challenge in this area has been summarised by M Friedman Ben-David[3] as: "how to assess progression towards defined outcomes, integrated abilities, different forms of medical knowledge, on-the-job learning, learning through assessment, error management, and portfolio evidence".

These new perspectives on assessment have been translated by the Danish National Board of Health into new guidelines for postgraduate medical education.[11, 12] In the future, residency programmes must address a broad spectrum of competencies in their objectives, learning strategies, and assessment. The broad spectrum of competencies included in the guidelines is derived from the seven roles and competencies defined by the CanMEDS 2000 Project. The guidelines stipulate that the objectives must represent an acceptable minimum standard of performance, the assessment must be criterion referenced, and it must be implemented as an on-going process rather than an end-of-training examination. Although the guidelines include suggestions for several assessment methods that may be used, they do not provide any further description, references or examples. Therefore, implementing the new guidelines poses a huge challenge, the more so because residency programmes in Denmark have not previously included specific assessment of trainees or specialist examinations. Until now, residency evaluation has consisted of at least three appraisal consultations of the trainee with a supervisor during each post and the trainee's evaluation of the programme.

This paper describes the intrinsic rational validation process, i.e. the reasons for particular decisions in test development, used in outlining an assessment programme for the first year of residency in anaesthesiology according to the new guidelines.[13] The description of that process may be helpful to educators and programme directors in meeting the challenges of the new paradigm. Also discussed is the applicability of the process to other residency programmes and higher-level training in anaesthesiology.

Method

A project supported by the Danish Society of Anaesthesiology and Intensive Care Medicine (DASAIM) was set up to develop the assessment programme for anaesthesiology residency year 1. The nationwide project was a joint venture with Copenhagen Hospital Corporation Postgraduate Medical Institute and was carried out by a working group of 20 people including members of DASAIM's board of education, consultants from various parts of the country and trainees. The authors of this article and the working group drafted an outline for the objectives and the assessment programme, based on a number of seminars, workshops and literature studies. The structure of the outline of the assessment programme was provided by five central questions with respect to residency assessment: Why, What, When, How and Who. Also taken into account was the programme's usability, defined as a function of the presumed validity, reliability, acceptability, effect on learning and costs as described by Van der Vleuten *et al.*[14] Finally, a key principle was to design an assessment programme that fitted in the context of daily practice in anaesthesiology departments and took account of the fact that trainees are perceived as part of the workforce. Between the seminars parts of the assessment programme were tried out in practice. On the basis of these experiences the assessment programme was continuously adjusted, especially its practicability and acceptability. All decisions were made by consensus and a draft was distributed to core faculty for review. The final curriculum, including the assessment programme, was implemented in all first year residency programmes in May 2001. The curriculum has been published in a book entitled 'The Education Book for 1st year training in Anaesthesiology'[15], which was sent out to all anaesthetists in Denmark. It is also available on the Internet, www.dasaim.dk.

Results

Examples of the key competences and objectives are listed in Table 1. The seven roles and competences have been reduced to six by integrating the role of 'health advocate' with that of 'medical expert', which includes identification of determinants of health and preventive measures that are important to risk and outcome in anaesthesia. The objectives include a list of procedures to be mastered by trainees at the end of the first year of training.

Table 1. Examples of key objectives in a first-year residency in anaesthesiology related to the new roles and competences

Role	*Objectives of some of the key competences*
1. Medical expert*	• Manage ASA group 1-2 patients for anaesthesia as a team leader and ASA group 3-5 as a team member • Manage simple intensive care patients • Manage simple emergencies including CPR as a team leader and initiate and sustain resuscitation of complicated patients and children until more experienced staff arrives • Manage procedures according to procedure list
2. Communicator	• Establish relations with patients and relatives characterised by understanding, respect, empathy and confidence. • Communicate effectively and encourage patient participation in decision making when appropriate • Recognize ethical and attitudinal dilemmas and discuss these with respect for patients' integrity, wishes and rights regarding information and participation in decisions
3. Collaborator	• Demonstrates effective communication with the multi-professional team • Work effectively as part of the team and demonstrate willingness and ability to take either the role as team leader or team member when appropriate • Respect the skills of other team members and their specific roles in different situations. Contribute to making the team appear homogenous and professional to the patients and relatives, and contribute to effective and consistent information to patients and relatives
4. Manager	• Is organised in his/her work, demonstrates planning, consideration, precaution, and responsibility. Works systematically and properly in a way that demonstrates a general overview of situations • Sets proper priorities among tasks and patients on the acute list according to patients' needs and the availability of staff and resources, and uses resources effectively
5. Scholar	• Can perform a critical analysis of a patient course using the literature and guidelines and reflect critically on own practice and that of the department in general • Is able to identify a clinical problem and reformulate it as a research question that can be answered through a literature search. Can perform a focused literature search, a critical appraisal of the literature, discuss the results and the implications for the specific clinical problem. Is able to present this in a short written report
6. Professional	• Recognize the importance of knowing own limitations, express this openly and call for timely assistance from more experienced colleagues when appropriate • Understands the importance of proper recording and registration of data and demonstrates this in written and oral communicationDemonstrates respect for effective patient management and management of the operating list. Demonstrates punctuality.

Note: *The list of objectives for this role is more elaborate than shown here.

Table 2. Assessment program for first-year residents in anaesthesiology: overview of the 21 individual instruments used.

	Competence	Roles addressed	When?	How?	Summative or formative
1.	Managing airways, intubation	• Medical expert	• Before on call in anaesthesia	• Observation by checklist + questions	• Summative
2.	Testing the anaesthesia machine	• Medical expert	• Before on call in anaesthesia	• Observation by checklist + questions	• Summative
3.	Principles of general anaesthesia	• Medical expert	• Before on call in anaesthesia	• Observation by checklist + questions	• Summative
4.	Principles of general anaesthesia in emergency patient	• Medical expert, health advocate	• Before on call in anaesthesia	• Observation by checklist + questions	• Summative
5.	Preoperative assessment of patient	• Medical expert, communicator, health advocate	• Before on call in anaesthesia	• Observation by checklist + questions	• Summative
6.	Advanced CPR	• Medical expert, communicator, collaborator, professional	• Before on call in anaesthesia	• Observation on mannequin, checklist + questions	• Summative
7.	Management plan for fluids and nutrition	• Medical expert	• Before on call in intensive care	• Audit on three patient plans + questions	• Summative
8.	Managing simple ventilator	• Medical expert	• Before on call in intensive care	• Observation by checklist + questions	• Summative
9.	Managing simple intensive care patient	• Medical expert, communicator	• Before on call in intensive care	• Observation by checklist + questions	• Summative
10.	Spinal anaesthesia	• Medical expert	• Whenever suitable	• Observation by checklist + questions	• Summative
11.	Epidural anaesthesia	• Medical expert	• Whenever suitable	• Observation by checklist + questions	• Summative
12.	Central venous catheter	• Medical expert	• Whenever suitable	• Observation of checklist + questions	• Summative

	Competence	Roles addressed	When?	How?	Summative or formative
13.	Registration of experience in anaesthesia and procedures	• Medical expert, scholar, professional	• 1st, 3rd, 6th, 9th, and 11th month	• Continuous registration, discussion with tutor	• Formative
14.	Registration of quality of spinal and epidural anaesthesia, central venous and arterial line	• Medical expert, scholar, professional	• 1st, 3rd, 6th, 9th, and 11th month	• Continuous registration, discussion with tutor	• Formative
15.	Patient communication skills	• Communicator, scholar	• After 6th month	• Patient survey, discussion with tutor	• Summative
16.	Team work and interpersonal skills	• Communicator, collaborator, manager, professional	• After 6th month	• Report from two supervisors according to checklist, discussion with tutor	• Summative
17.	Managing operating list on call	• Collaborator, manager, professional	• After 6th month	• Assessment by senior colleague according to checklist and questions to the team.	• Summative
18.	Anaesthesia and complicating medical diseases	• Medical expert, scholar	• Whenever suitable	• Written assignment, assessment by supervisor	• Summative
19.	Patient management plan, a critical appraisal of practice	• Medical expert, health advocate, scholar	• Whenever suitable	• Written assignment, assessment by supervisor	• Summative
20.	Evidence-based medicine exercise	• Medical expert, scholar	• Whenever suitable	• Written assignment, assessment by supervisor	• Summative
21.	Using the learning portfolio	• Scholar	• 1st, 3rd, 6th, 9th, and 11th month	• Discussion with tutor	• Formative

'The Education Book' contains a list of key textbooks and other references relevant to first year residents. Each anaesthesiology department designs it's own training programme targeted at the stated objectives. In addition, the departments in each of the three geographical regions in Denmark, North, South and East, collaborates on designing 10 one-day seminars to be attended by the residents. Issues of basic scientific knowledge and anaesthesia knowledge are addressed in the training programmes and the seminars. The resulting assessment programme consists of 21 specific tests or assessment strategies (Table 2). A few examples of the assessment formats are given in the Appendix. The programme is presented below under the headings why, what, when, how and who.

Why?

The basic question to be answered before an assessment programme can be designed asks about the purpose of the programme. The working group identified several purposes of assessment of first year residents. Assessment assures trainees and stakeholders that minimum standards are met. The assessment programme also serves as an aid to the educational programme. One of the main purposes is to reduce trainees' stress levels in emergency situations by increasing their confidence in their own abilities. Another purpose is to take advantage of the fact that assessment drives learning and that trainees can learn through assessment. The assessment programme is also aimed at fostering a positive attitude towards systematic collection of data and information on quality of care and a continuous and systematic review of and reflection on clinical practice. Finally, assessment is used for a summative purpose that is the selection of trainees for higher-level residency training in anaesthesiology. The assessment programme was constructed to meet all these purposes.

What?

This question refers to the content of assessment. What do we need to evaluate to ensure that assessment fulfils the purposes we have identified? Table 2 lists the 21 competencies that comprise the assessment programme. The assessment programme should reflect the objectives including the wide spectrum of roles and competences. Being aware that only a sample of the learning of residents can be tested, the group aimed at constructing a programme focused on competences and procedures that were perceived to be of major importance or more complex. The first twelve competences are related to procedural skills and standard patient management skills. It was decided to assess managing airways and intubation skills because of the potential risk of life threatening complications.[16] Checking the anaesthetic machine was selected because this is often neglected by physicians,[17] and basic life support skills because an-

aesthesiology residents have been shown to be unable to perform these skills properly even while in training and without recognizing this themselves.[18] Skills such as spinal and epidural anaesthesia and central venous line were included because of uncertainties regarding the learning curves and the relation between experience and complications.[19, 20] In item 14 these skills are assessed from the perspective of other roles. Other procedures such as laryngeal mask and peripheral venous catheter were perceived as more simple and therefore not requiring the same attention. Item 13 consists of continuous documentation of trainees' experiences during the year. A special log (item 14) tests four procedures: spinal and epidural anaesthesia, central venous line, and arterial catheter. Patient communication skills (item 15) are assessed by a patient survey and items 16 and 17 address managerial, collaboration, and professional competences. Items 18-20 assess reflection on practice and academic competence in three areas. Item 18, 'anaesthesia and complicating diseases' evaluates the choice of anaesthetic practice for a standard case and changes of practice related to various complicating conditions, item 19 requires the trainee to write a management plan and reflections on the actual perioperative course, and item 20 addresses the trainee's performance in applying the principles of evidence-based-medicine. Finally, the trainees' competence in self-directed learning is assessed (item no 21).

How?

One of the aims of the assessment programme is to facilitate learning through assessment. Hence both trainees and assessors are informed of the assessment programme and its content. If a trainee fails a test, the general rule is that proper remediation is instituted and the assessment procedure repeated. A handbook for assessors contains guidelines on how to handle situations where trainees do not pass a test and other potential problems, for instance difficulties arising from the conflicting roles of preceptor and assessor being combined in one person at the same time.

How the 21 competencies are assessed is shown in Table 2. For the first twelve competences direct observation is used with scoring on a checklist, which also includes some standard questions on, for example indications, contra-indications, risk of complications, their treatment and preventive measures, drug choice and dose, ethical considerations, etc. The scoring on the checklist is binary, yes/no, and all items on the checklist must be correct for the performance of the skill to be approved. These tests are summative and at the 'shows how' level. They must be passed before the trainee is allowed to perform without close supervision or assigned to be on call.

The assessment method used in item 13 is documentation of the trainee's experiences over the course of the year. The data to be entered

are related to patient cases, procedures and complications. This method tests at the 'does' level and is used for formative assessment. The database is reviewed regularly with the trainee's tutor and helps direct the trainee's activities or indicates when remediation is needed. At the same time the database serves as documentation of whether objectives of procedural skills have been met.

The four procedures of item 14 are tested by means of a log, which monitors the quality of the procedures, by use of the Cusum technique.[21] The principle in the Cusum technique is that every time the procedure is performed, the trainee records success or failure by a specific scoring system, which takes the failure rate of experienced anaesthetists into account. This instrument tests at the 'does' level and over time the scoring provides a learning curve identifying trainees in need of remediation.

Patient communication skills (item 15) are assessed by a patient survey including 10 questions and 25 patients. The questionnaire is predesigned and included in 'The Education Book'. The trainee is responsible for carrying out the survey and for reporting the results to the tutor. This is also a test at the 'does' level. If the results of the survey do not meet the set criterion, the trainee will be observed in at least five patient encounters. Performance is scored on a checklist that is similar to the questions in the survey.

Managerial, collaboration and professional competences (items 16 and 17) are assessed by two individual senior staff members for daytime performance and by one senior on call. The competences are scored on specified rating scales. For the on call assessment the rating scales include questions about the organisation and function of the anaesthetic team and the hospital organisation, and about procedures in emergencies and catastrophes.

Reflection on practice and academic competence (items 18-20) are assessed by three written assignments. 'Anaesthesia and complicating diseases' requires the trainee to discuss the choice of anaesthetic practice for a standard case and changes of practice related to various complicating conditions. The written assignment on the patient management plan (item 19) requires the resident to describe a management plan for a patient of his or her own choice, followed by reflections on the actual perioperative course. A third written assignment (item 20) is an exercise in evidence-based-medicine relating to a case selected by the trainee. Supervisors assess the assignments using standard review formats.

Finally, the trainees' competence in self-directed learning (item 21) is assessed by use of a learning portfolio. An outline and examples of the principles of the learning portfolio are specified to trainees and tutors. The learning portfolio is used for formative evaluation and as subject of discussion at regular meetings of tutor and resident.

When?
The timing of the individual tests is in principle left to the initiative of the trainees. A resident has to notify a senior staff member that he or she feels ready to be assessed. The assessment programme is designed to ensure that individual tests are timed in accordance with the aim of the assessment and the competence at stake. Thus assessment of procedural skills is mainly placed at the beginning of the year and assessment of academic competencies and reflection on practice in the second half of the training year, when the trainees have gained some experience in various procedures and patients. Assessment at the 'does level' through continuous collection of data and assessment of self-directed learning abilities is an ongoing activity throughout the year.

Who?
The assessment programme is designed in such detail that any specialist in anaesthesiology can function as an assessor. In each department one consultant is responsible for resident education and each resident is assigned a tutor, who sees to it that tests are taken in due time and that the trainee returns the signed forms to document satisfactory performance of competencies. The consultant responsible for education approves by signature a final list of the summative assessments.

Discussion

This paper describes how an assessment programme was designed according to the new paradigm that is facing residency education. The assessment programme does not include a national, uniform end-of-training examination. It is entirely based on timely assessment in clinical practice in the individual departments. Nevertheless the programme is implemented nationwide and the detailed outline of the assessment programme, which includes the checklists to be used, ensures that all trainees are assessed by the same methods and standards. The programme has several advantages. Its independence of time and place is an advantage in a system where not all trainee posts in all hospitals start at the same time. Furthermore, the programme is proactive in that trainees' performance on procedures is assessed at the time when they are actually expected to perform the procedures in clinical practice, which seems logical from a quality of care point of view. This approach may speed up the learning process as was demonstrated by Long in a neurosurgery residency.[22] Another advantage is the opportunity for timely remediation. Finally, as expressed by working group members who piloted the indi-

vidual tests, "one positive side effect is that as an assessor you have to keep up to the standards". The disadvantages that were foreseen are the conflict inherent in combining the roles of appraiser and assessor in the same person and the problem of educating the assessors.[9]

Any assessment programme will have to meet requirements of reliability, validity, feasibility, acceptability, and effect on learning and we are currently investigating these issues.[23, 24] In the following we discuss these issues and how they have influenced our choice of design and methods.

Regarding the reliability of observations in residency assessment Noel *et al*[25] revealed serious problems in internal medicine and showed that faculty in university hospitals were more accurate than staff in community hospitals. They demonstrated that accuracy was improved by the use of checklists. Whether the content of the items in our assessment programme is sufficiently detailed in this respect remains to be shown. The reliability of logbook data and learning portfolios is problematic[26, 27] and we are currently using these methods for formative assessment only. Patient ratings of communication skills from 20-40 patients on 5 to 10 items have been shown to be sufficient for obtaining a reproducible, meaningful indicator of residents' skills.[28]

In order to be valid the assessment programme should reflect the stated objectives including the wide spectrum of roles and competences. Being aware that any assessment programme can only test a sample from the entire domain, the group aimed to construct a programme that focused on competences and procedures that were considered to be of major importance or more complex.

One important question is whether introducing this assessment system compared to the previous system, which has with no assessment, will result in more competent residents by the end of the first year of anaesthesiology residency. There are reasons to believe that introducing assessment will have a positive effect on learning as well as on teaching in the clinical setting.[29] In dealing with many concerns regarding feasibility, acceptability, and cost, the working group worked hard to narrow down the programme. The somewhat limited number of observed tests may have a negative effect on learning and limit trainees' experiences as a result of the steering effect of assessment on trainees' behaviour.[30] This negative effect may be counterbalanced, however, by the positive effect of the continuous log on behaviour, which it is hoped will direct training to the entire set of objectives. Hunskaar & Seim[31] demonstrated such an effect when they introduced a list of objectives for procedures in an undergraduate curriculum without any other educational intervention. The positive as well as negative effects on learning will have to be studied

further and the implementation of this assessment programme will be followed by an extensive evaluation.

An important question about an assessment programme is whether there are better ways to do the same thing.[32] A final examination including an OSCE is a well-described alternative.[33] A knowledge-of-skills test has been shown to correlate well with performance-based testing, and applied as a progress test would emphasise the effect on learning.[34] Nevertheless, this kind of test cannot replace observation in practice and would not be appropriate for testing the broad spectrum of competences as required by the new paradigm.

Hays & Wellard[35] presented a model of an in-training assessment programme in postgraduate training for general practice. The main difference from our programme is that theirs is aimed at formative assessment and that the summative part is a final certification examination.[9] Hays & Wellard[35] emphasised the problem of designing an assessment programme that is congruent with the values and tradition of the speciality. Family Medicine in Australia traditionally has a strong focus on adult learning principles and self-directed learning with feedback being regarded as a vital part of these principles. This means that formative in-training assessment is acceptable, whereas summative in-training assessment is perceived as incongruent with these principles. Interestingly, summative assessment is accepted in the form of a certificate examination. End-of-training examinations are not highly valued in some parts of Europe. Karle & Nystrup[36] summarised how pedagogical, methodological, legal and resource problems influence these aspects. They proposed a better alternative a comprehensive formative evaluation involving both trainees and institutions.

One of the controversial elements of the programme described in this article is that it does not test cognitive knowledge to the extent that traditional board examinations do.[37] Board examinations include a comprehensive test of knowledge and they are generally perceived as the 'gold standard'. Slogoff *et al.*[38] studied the relation between good clinical performance and results on a board examination in anaesthesiology. They found a strong relation between clinical skills ratings and certification success rates. Clinical skills ratings consisted of the consultant's judgement whether he/she would permit the anaesthetist to provide anaesthesia care for him/her in three increasingly complex scenarios. Reich *et al.*[39] demonstrated that tests in the academic domain administered within the first month of anaesthesiology residency were mildly predictive of clinical performance whereas tests administered after 6 and 13 months were not. McLeskey & Ward[40] found that senior residents appeared to perform worse on written examinations compared to junior residents with less than 2 years of training. The question is whether there is a causal rela-

tionship between board certification and good clinical performance or whether it is a matter of co-existing phenomena. Friedman Ben-David[3] addressed the question of the relation between practice and knowledge in a paper on the future role of assessment and stated that: "Core knowledge may take on a completely different meaning when decisions are made as to what forms of knowledge should be separated from action and which forms of knowledge should be inferred from action". According to Rasmussen[41] the first step in practice learning is the acquisition of skills, and this can be achieved before the full theoretical knowledge for their application has been acquired.[41, 42] As skills are acquired, trainees learn to follow rules that constitute appropriate responses to most situations most of the time. At higher levels solutions are derived from broad experience and not directly related exclusively to the specific information that is available in the current situation. Our assessment programme is designed for first-year anaesthesiology residency training, and assessment of procedural skills by observation and checklists corresponds to the lower levels of practice learning, whereas the written assignments and the learning portfolio target the more experienced level in the latter half of the year. In that sense the design of our programme seems appropriate for first year residents.

One aim of our assessment programme is to decide whether trainees can move on to higher-level residency training. An important issue in this respect is to assess whether trainees are able to perform basic tasks relevant to anaesthesiology by testing the 'shows how' level. It is also important to test whether the trainees are able to accommodate to the speciality's standards of professionalism. Slogoff *et al*[38] found that consultants' indications of trainees' professional competence were strongly linked to clinical skills ratings. Cusimano *et al*[43] showed in neurosurgery residencies that reasons for dismissals were issues of professionalism and not cognitive deficits. Also in that sense it appears rational that our assessment programme should include aspects of professionalism.

The design and methods used in our assessment programme may not be suitable for higher levels of training or other specialities. The use of checklists, for example, has been shown to be inappropriate in higher levels of experience and more complex skills.[44] In anaesthesiology high-stakes simulators are likely to be used in the future for testing complex skills and integrated abilities.[33]

Different specialities have different values, traits and preferred learning styles and the education and assessment strategies will have to be in keeping with those.[45-47] Baker *et al*. [45] identified a predominance of the 'Accommodator learning style' emphasising clinical experience in anaesthesiology and the authors point to the importance of assignments in non-dominant areas such as conceptual knowledge, writing and research.

In our assessment programme the academic activities in the three written assignments are in accordance with this perspective. The emphasis on various domains may be different for other specialities.

Conclusion

We believe that the process and considerations involved in outlining the assessment programme for the first year of anaesthesiology residency training may serve as a template for those who wish to design similar programmes for other specialities. The specific design and the methods will probably have to be adapted if they are to be used in higher levels of training in anaesthesiology and in other specialities.

Acknowledgements: The working group that compiled 'The education book for first year residency in anaesthesiology' included: Andersen N, Berlac PA, Bested K, Callesen T, Christensen P, Jensen E, Jensen JW, Lemholt K, Lund J, Malling B, Mandøe H, Nørregaard O, Pedersen BD, Petersen JA, Ravlo O, Ravn L, Ringsted C, Skjelsager K, Sprehn M, Østergaard D.

References

1. CanMEDS 2000: Extract from the CanMEDS 2000 Project Societal Needs Working Group Report. Med Teach 2000; 22: 549-54.
2. Accreditation Council for Graduate Medical Education. Outcome project (http://www.ACGME.org), 2000.
3. Friedman Ben-David M. The role of assessment in expanding professional horizons. Med Teach 2000; 22: 472-7.
4. Miller G. Conference summary. Acad Med 1993; 68: 471-4.
5. Dauphinee WD. Assessing clinical performance. Where do we stand and what might we expect? JAMA 1995; 274: 741-3.
6. Harden R, Gleeson F. Assessment of clinical competence using an objective structured clinical examination (OSCE). Med Educ 1979; 13: 41-54.
7. Van der Vleuten CPM, Swanson DB. Assessment of clinical skills with standardized patients: state of the art. Teach Learn Med 1990; 2: 58-76.
8. Gray JD. Global rating scales in residency education. Acad Med 1996;71:S55-63.
9. Feletti G, Cameron D, Dawson-Saunders B, Des Groseilliers J, Dooley B, Farmer E, McAvoy P. In-training assessment. In: Newble D, Jolly B, Wakeford R, eds. The certification and recertification of doctors: Issues in the assessment of clinical competence. Cambridge University Press. 1994: 151-66.
10. Shepard LA. The role of assessment in a learning culture. Educ Research 2000; 29: 4-14.
11. Ministry of Health. The Future specialist. Report from the Specialist Commission. Report no. 1384. Copenhagen: Statens Information, 2000.
12. National Board of Health. Guidelines for Curriculum of Specialist Education. Copenhagen: National Board of Health, 2001.

13. Ebel RL, Frisbie DA. Essentials of educational measurement. In: Ebel RL, Frisbie DA, eds. 5th edn. Englewood Cliffs, New Jersey: Prentice Hall, 1991: 102-5.
14. Van der Vleuten CPM, Scherpbier AJJA, Dolmans DHJM, Schuwirth LWT, Verwijnen GM, Wolfhagen HAP. Clerkship assessment assessed. Med Teach 2000; 22: 592-600.
15. Ringsted C, Østergaard D. The education book for 1st year training in anaesthesiology [Uddannelsesbog. Introduktionsuddannelsen i anæstesiologi, H:SPMI] (http://www.dasaim.dk and http://www.hspmi.dk), Copenhagen, Denmark: CHC Postgraduate Medical Institute, 2001.
16. Crosby ET, Cooper RM, Douglas MJ, Doyle DJ, Hung OR, Labrecque P, Muir H, Murphy MF, Preston RP, Rose DK, Roy L. The unanticipated difficult airway with recommendations for management. Can J of Anaesth 1998; 45: 757-76.
17. Klopfenstein CE, van Gessel E, Forster A. Checking the anaesthetic machine: self-reported assessment in a university hospital. Eur J of Anaesth 1998; 15: 314-9.
18. Noordergraaf GJ, Be WK, Sabbe M, Diets RF, Noordergraaf A, van Hemelrijck J. Training needs and qualifications of anaesthesiologists not exposed to ALS. Resuscitation 1999; 40: 147-60.
19. Kopacz DJ, Neal JM, Pollock J. The regional anaesthesia 'learning curve'. What is the minimum number of epidural and spinal blocks to reach consistency? Regional Anesthesia 1996;21:182-90.
20. Konrad C, Scüpfer G, Wietlishbach M, Gerber H. Learning manual skills in anesthesiology. Is there a recommended number of cases for anesthestic procedures? Anesth Analg 1998; 86: 635-9.
21. Kestin IG. A statistical approach to measuring the competence of anaesthetic trainees at practical procedures. Br J Anaesth1995; 75: 805-9.
22. Long DM. Competency-based residency training: the next advance in graduate medical education. Acad Med 2000; 75: 1178-83.
23. Van der Vleuten CPM. The assessment of professional competence: developments, research and practical implications. Adv Health Sci Educ 1996; 1: 41-67.
24. Fowell SL, Southgate LJ, Bligh JG. Evaluating assessment: the missing link? Med Educ 1999; 33: 276-81.
25. Noel GL, Herbers JE, Caplow MP, Cooper GS, Pangaro JN, Harvey J. How well do internal medicine faculty members evaluate the clinical skills of residents? Ann Int Med 1992; 117: 757-65.
26. Pitts J, Coles C, Thomas P. Enhancing reliability in portfolio assessment. 'shaping' the portfolio. Med Teach 2001: 23: 351-6.
27. Raghoebar-Krieger HMJ, Sleiffer D, Bender W, Stewart RE, Popping R. The reliability of logbook data of medical students: an estimation of interobserver agreement, sensitivity and specificity. Med Educ 2001; 35: 624-31.
28. Swanson, DB, Webster GD, Norcini JJ. Precision of patient ratings of residents' humanistic qualities: how many items and patients are enough? In: Bender W, Hiemstra RJ, Scherpbier AJJA, Zwierstra RP, (Eds.) Teaching and Assessing Clinical Competence. Groningen: BoekWerk Publications, 1990: 424-31.
29. Petrusa, ER. Clinical performance assessments. In Norman GR, Van der Vleuten CPM, Newble DI, (eds.) International Handbook of Research in Medical Education. Dordrecht: Kluwer Academic Publishers, 2002: 673-709.
30. Wakeford RE, Southgate L. Postgraduate medical education. Changing trainees' study approaches by changing the examination. Teach Learn Med 1992; 4: 210-3.
31. Hunskaar A, Seim AH. The effect of a checklist on medical students' exposures to practical skills. Med Educ 1984; 18: 439-42.
32. Schriven, M. The nature of evaluation part I: Relation to psychology. Practical Assessment, Research & Evaluation, 6(11). (http://ericae.net/pare/getvn.asp? v=6 &n=11), 1999.
33. Siker ES. Assessment of clinical competence. Current Opinion in Anaesthesiology, 1999; 2: 677-84.

34. Ram P, van der Vleuten C, Rethans JJ, Schouten B, Hobma S, Grol R. Assessment in general practice: the predictive value of written-knowledge tests and a multiple-station examination for actual medical performance in daily practice. Med Educ 1999; 33:197-203.
35. Hays R, Wellard R. In-training assessment in postgraduate training for general practice. Med Educ 1998; 32: 507-13.
36. Karle H, Nystrup J. Comprehensive evaluation of specialist training: an alternative to Board examinations in Europe? Med Educ 1995; 29: 308-16.
37. Zorab JSM. The educational activities of the European Academy of Anaesthesiology. Minerva Anestesiologica 1998; 64: 477-82.
38. Slogoff S, Hughes FP, Hug CC, Longnecker DE, Saidman LJ. A demonstration of validity for certification by the American Board of Anesthesiology. Acad Med 1994;69:740-6.
39. Reich DL, Uysal S, Bodian CA, Gabriele S, Hibbard M, Gordon W, Sliwinki M, Kayne RD. The relationship of cognitive, personality, and academic measure to anesthesiology resident clinical performance. Anesth Analg 1999; 88: 1092-1100.
40. McLeskey CH, Ward RJ. Validity of written examination. Anesthesiology 1978; 49: 224.
41. Rasmussen J. Skills, rules, knowledge, signals, signs and symbols and other distinctions in human performance models. IEEE Transactions Systems, Man and Cybernetics 1983; 123: 257-66.
42. Rasmussen J, Lind M. A model of human decision making in complex systems and its use for design of system control strategies. Riso National Laboratory Report. (Riso-M-2349, Roskilde, Denmark), 1982.
43. Cusimano MD, Yonke AM, Tucker WS. An analysis of attrition from Canadian neurosurgery residency programs. Acad Med 1999; 74: 925-31.
44. Hodges B, Regehr G, McNaughton N, Tiberius R, Hanson M. OSCE checklists do not capture increasing levels of expertise. Acad Med 1999: 74: 1129-34.
45. Baker JD, Cooke JE, Contoy JM, Bromley HR, Hollon MF, Alpert CC. Beyond career choice: the role of learning style analysis in residency training. Med Educ 1988;22:527-32.
46. Merril JM, Camacho Z, Laux LF, Lorimor B, Thornby JI, Vallbona C. Uncertainties and ambiguities: measuring how medical students cope. Med Educ 1994; 28: 316-22.
47. Stilwell NA, Wallick MM, Thal SE, Burleson JA. Myers-Briggs type and medical specialty choice: a new look at an old question. Teach Learn Med, 2000; 12: 14-20.

Appendix: Examples of checklists/assessment forms

Example of checklist used for observing a procedure - conduct of a spinal anaesthesia

The resident indicates when he/she is ready for assessment and finds a senior colleague that can supervise him/her according to this checklist. All items on the checklist should be correct in order to approve the competence.

	Yes	No
• Outlines plan for anaesthesia including description of indication of proper effect, acceptable blood pressure values, plan for managing side-effects or unwanted effects	☐	☐
• Makes proper preparations in the anaesthetic and operating rooms	☐	☐
• Can explain choice of anaesthetic and dose related to the planned surgery	☐	☐
• Informs and instructs patient properly regarding the procedure	☐	☐
• Performs proper sterile preparation; check medicine and instruments	☐	☐
• Makes correct positioning of patient in collaboration with assistant	☐	☐
• Indicates correct site of needle prick	☐	☐
• Uses proper sterile technique	☐	☐
• Uses correct needle technique and assures correct needle position before injecting anaesthetic	☐	☐
• Performs proper observation and monitoring of patient after injection of local anaesthetic	☐	☐
• Initiates proper measures to support cardiovascular function	☐	☐
• Makes proper test of the analgesia and indicate segmental level correctly	☐	☐
• Can name indications and contra-indication to spinal anaesthesia	☐	☐
• Can name at least three complications and risks of spinal anaesthesia; can describe precautions regarding prevention and describe proper management of these complications.	☐	☐
This competence is acceptable	☐	☐

Name of supervisor: Name of resident: Date:

Example of assessment form used for assessing organisational skills and collaboration in the anaesthetic department.

Two senior colleagues independently assess this competence using this form. On behalf of these two assessments the tutor makes an overall assessment. Mean score from the two seniors should be ≥3 in order to approve the competence

	Don't know	Poor	Fair	Good	Very good
Organizing own work		1	2	3	4
• Demonstrates proper planning and order of priority of work	☐	☐	☐	☐	☐
• Shows thoughtfulness and vigilance during work and management of the situations	☐	☐	☐	☐	☐
• Demonstrates tidiness and structured approach to practise	☐	☐	☐	☐	☐
• Demonstrates punctuality in appearance and gives information on whereabouts and how he/she can be contacted or called	☐	☐	☐	☐	☐
Professional appearance					
• Knows his/her own limits of competence	☐	☐	☐	☐	☐
• Demonstrates responsibility towards work assignments	☐	☐	☐	☐	☐
• Demonstrates ability to critical assessment of quality of practice	☐	☐	☐	☐	☐
• Demonstrates ability to handle critical incidents and mistakes	☐	☐	☐	☐	☐
• Shows respect towards demands for effective management of the operating list	☐	☐	☐	☐	☐
Team collaboration and communication					
• Works effectively in a team and assumes equally well the role as team member and team leader when appropriate	☐	☐	☐	☐	☐
• Demonstrates understanding and respect for other team members, their professional competence and situation-specific roles and assignments	☐	☐	☐	☐	☐
• Is receptive and open, respects other people's opinion and contributes with own expertise when appropriate	☐	☐	☐	☐	☐

Overall assessment by the tutor on behalf of the scoring of two assessors	Yes	No
The resident has been assessed by two senior colleagues:	☐	☐
Organization of own work is acceptable	☐	☐
Professional appearance is acceptable	☐	☐
Collaboration with the team is acceptable	☐	☐

Name of tutor: Name of resident Date:

Example of assessment: professional development and the use of the learning portfolio

This form is used at the regular meetings between tutor and resident

	Yes	No
• Has the resident followed the learning plan from the last meeting?	☐	☐
• Has the objectives been met, i.e. can the resident describe what he/she has learned?	☐	☐
• If the objective has not been met – can the resident give a reasonable explanation?	☐	☐
• Has the learning portfolio been useful to the resident?	☐	☐
• Has the resident any suggestions regarding things that could be done differently in planning the next period?	☐	☐
• Has the mandatory assessments been completed according to the plan?	☐	☐
• Has the resident completed the logbook of experience and the Cusum scoring?	☐	☐
• Does the logbook shows proper breadth in experience?	☐	☐
• Does the Cusum scoring shows proper development in the resident routine regarding the procedures?	☐	☐
• Has the resident accomplished the issues in the programme appropriately according to this stage of his/her education?	☐	☐
The professional development is acceptable	☐	☐

Name tutor: Name resident: Date:

CHAPTER 2

Content validity of ITA

Published in Acta Anaesthesiologica Scandinavica 2002; 46: 1119-23

Consultants' opinion on a new practice based assessment programme for first year residents in anaesthesiology.

C Ringsted, D Østergaard, and A. Scherpbier

Abstract

Background: Assessment in postgraduate education is moving towards using a broad spectrum of practice-based assessment methods. This approach was recently introduced in first-year residency in anaesthesiology in Denmark. The new assessment programme covers: clinical skills, communication skills, organisational skills and collaborative skills, scholarly proficiencies, and professionalism. Eighteen out of a total of 21 assessment instruments are used for pass/fail decisions. The aim of this study was to survey consultants' opinion of the programme in terms of the representativeness of competencies tested, the suitability of the programme as a basis for pass/fail decisions and the relevance and sufficiency of the content of the different assessment instruments.
Methods: A description of the assessment programme and a questionnaire were sent to all consultants in anaesthesiology in Denmark. The questionnaire consisted of items, to be answered on a five-point scale, asking the consultants' opinions about representativeness, suitability, and content of the programme.
Results: The response rate was 251/382 (66%). More than 75% of the respondents agreed that the assessment programme offered adequate coverage of the competencies of a first-year resident and was appropriate for making pass/fail decisions. There was strong agreement that the content of the 18 tests used for pass/fail decisions was relevant and sufficient for pass/fail decisions.
Conclusion: Judging from the consultants' opinions, the assessment programme for the first-year residency in anaesthesiology appears to be appropriate regarding the range of competencies assessed, the appropriateness as a basis for pass/fail decisions, and regarding of the content of tests used for pass/fail decisions. Further studies are needed to assess the feasibility and acceptability of the programme in practice.

Introduction

Assessment in postgraduate medical education is changing regarding content, methods and concept. A broader spectrum of knowledge and skills are required and this must be reflected in the assessment.[1, 2] There is a move towards assessment by observation of residents' performance in real practice and collection of information about performance from different sources.[1, 3] Furthermore, the concept of assessment as an end of training activity is increasingly being replaced by the view that assessment is an integral part of the learning process.[4]

These ideas are reflected in the new guidelines for specialist training issued by the Danish National Board of Health.[5] According to the guidelines, the objectives of residency programmes must cover a broad spectrum of competencies. Furthermore the new guidelines request the specialties to outline a formal assessment programme that must be implemented as an on-going process rather than an end-of-training examination. Although the guidelines include suggestions for several assessment methods that may be used, they do not provide any further description, references or examples.

A new curriculum for the first year of residency training in anaesthesiology was developed by a group of anaesthetists according to the new guidelines and introduced in the spring of 2001.[6] The assessment programme is practice-based and consists of 21 different instruments and methods, with tests being spread over the year. Implementing the new programme poses a huge challenge, the more so because residency programmes in Denmark have not previously included formal clinical assessment of residents or specialist examinations. Until now, residency evaluation has consisted of three appraisal consultations of the resident with a supervisor during each post, and the resident's evaluation of the programme.

Our assessment programme will be used by consultants nation wide. Specialist training in anaesthesiology in Denmark lasts 5 years of which 6 months are supplementary training in another speciality relevant to anaesthesiology. The first and final year of residency training can take place in general hospitals, whereas years 2 and 3 are performed in highly specialized university hospitals. The aim of this study was to survey consultants' opinion of the new assessment programme in terms of the representativeness of competencies tested, the suitability of the programme as a basis for pass/fail decisions and the relevance and sufficiency of the content of the different assessment instruments. At the same time we wanted to investigate if there was a relationship between the consultants' opinions and their having teaching responsibilities towards first-year versus second- and third-year residents.

Method

Assessment programme

The assessment programme comprises 21 different assessments in which various instruments and methods are used. An overview of the programme is presented in Appendix 1. It lists for each of the 21 assessments: the competency to be assessed, when the assessment is to be scheduled, the method used, who performs the assessment, and whether

it is summative or formative. Summative assessment means that the results are used for pass/fail decisions and formative assessment aims at providing feedback to the resident to guide further learning.

Eighteen of the 21 assessments are used for pass/fail decisions. Of these 18 assessments, 11 tests addressed procedural skills. Patient management plans, communication skills, collaborative skills and managerial skills were each assessed by one test, and three assessments dealt with theoretical reflection on practice. Observation in practice with scoring on detailed checklists was the method used for the assessment of procedural skills. Patient management plans regarding fluids and nutrition were assessed by a review of the resident's patient records. A patient survey was used to assess the resident's communication skills. Supervisors rated collaborative and managerial skills using a detailed checklist. There were three written assignments, which assessed theoretical reflection on practice. Items in the various instruments were scored dichotomously: yes or no. Residents had to obtain a pass on all test items. There were three instruments used for formative purposes: two logbooks in which the resident's clinical experiences with patients and procedures were recorded, and a learning portfolio for individual learning plans and reflection on learning. These instruments were discussed during regular meetings of resident and tutor.

Consultants' opinions

A questionnaire was sent out to all consultant anaesthetists in Denmark together with the curriculum and the assessment programme. The respondents were asked whether in their opinion 1) the range of competences of the assessment programme was representative of the competencies of first-year residents, 2) the assessment programme was appropriate for reaching a pass/fail decision about the first year, and 3) the assessment programme was appropriate to determine if the resident is ready to move on to the second year. For each of the 18 tests used for summative purposes the respondents were asked to indicate whether 4) the competency addressed was important 5) the content of the test was relevant, and 6) the content of the test was sufficient for an adequate assessment of competence. The respondents were asked whether they had first or second/third-year residents in their department. Finally they were invited to add comments.

Respondents were asked to indicate their answers to questions 1-6 on a five-point scale, where 1=not at all and 5=to a high degree. A score at or above 4 from over 75% of respondents was considered an indication of strong agreement among consultants. Results are reported as medians (25-75 quartiles). Data were analysed using median test for comparison between groups (SPSS 10.0).

Results

The questionnaires were sent out to 382 consultants and 251 (66%) responded. The responses showed that 89 (36%) consultants had first-year residents in their departments, 114 (45%) had second/third year and occasionally first year residents, and the remaining 48 (19%) had no residents in years 1-3 in their department.

The questions about the scope of the tests being representative of first year competence (question 1) and the appropriateness of the test as a basis for pass/fail decisions (questions 2 and 3) yielded scores of 4 or 5 from more than 75% of the respondents (Table 1). There were no statistically significant differences in responses between the groups of consultants.

Table 1. Consultants' median ratings on a scale from 1=not at all to 5=to a high degree in response to questions about the assessment programme. The 25-75 percentiles are given in brackets.

Questions	Median (25-75 percentiles)
Is the range of competencies in the assessment programme representative of the competencies expected of a first year resident?	4 (4-5)
Is the assessment programme appropriate as a basis for pass/fail decisions regarding the first-year of residency training?	4 (4-5)
Is the assessment programme appropriate for decisions regarding residents' competency to move to the second year of residency training?	4 (4-5)

Responses to the questions about each of the 18 tests used for pass/fail decisions are shown in Table 2. The median of the responses to the questions about the importance of the competency reflected in the test was 4 or 5 for each of the 18 tests. The lower quartile was 4 in 16 of the tests and 3 in two of the tests. The latter two were *'Managing operating list on call'* (23% of responses = 3; 3% < 3) and *'Evidence-based-medicine exercise'* (26% of responses = 3, 7% < 3). The scores of the consultants with first-year residents in their departments were significantly lower (median 4) (3-4) compared with those of consultants with second and third year residents: median 4 (3-5), P=0.000.

The questions about the relevance and sufficiency of test content yielded median responses of 4 or 5 for all tests with the lower quartile always at or above 4. There were no significant differences between the groups of consultants in their answers to these two questions.

Table 2. Consultants (n=251) median responses on a scale from 1=not at all to 5=to a high degree to questions regarding the content of 18 tests used for pass/fail decisions. In brackets are the 25 and 75 percentiles.

	How important is this competence?	Is the content of this test relevant?	Is the content of this test sufficient?
Managing airways and intubation	5 (5-5)	5 (5-5)	4 (4-5)
Testing anaesthesia machine	5 (5-5)	5 (4-5)	4 (4-5)
Principles of standard general anaesthesia	5 (5-5)	5 (5-5)	5 (4-5)
Principles of standard general anaesthesia in emergency patient	5 (5-5)	5 (4-5)	5 (4-5)
Preoperative assessment of patient	5 (5-5)	5(4.25-5)	5 (4-5)
Advanced CPR skills	5 (5-5)	5 (5-5)	5 (4-5)
Management plan for fluids and nutrition	4 (4-5)	5 (4-5)	5 (4-5)
Managing simple ventilator	5 (4-5)	5 (4-5)	5 (4-5)
Managing simple intensive care patient	4 (4-5)	5 (4-5)	5 (4-5)
Spinal anaesthesia	5 (4-5)	5 (5-5)	5 (4-5)
Epidural anaesthesia	4 (4-5)	5 (4-5)	5 (4-5)
Central venous catheter	4 (4-5)	5 (4-5)	5 (4-5)
Patient communication skills	5 (4-5)	5 (4-5)	5 (4-5)
Team work and interpersonal skills	5 (4-5)	5 (4-5)	4 (4-5)
Managing operating list on call	4 (3-5)	5 (4-5)	5 (4-5)
Anaesthesia and complicating medical diseases: Written assignment	4 (4-5)	5 (4-5)	4 (4-5)
Patient management plan: Written assignment	4 (4-5)	5 (4-5)	5 (4-5)
Evidence-based medicine exercise: Written assignment	4 (3-5)*	4 (4-5)	5 (4-5)

*=Consultants from departments with first-year residents scored significantly lower compared with their colleagues who did not have teaching responsibilities towards first-year residents, P=0.000.

One hundred and twenty-eight consultants commented on the assessment programme. In summary, 58 consultants praised the initiative. They used words like 'excellent' or 'brilliant' and thought that the assessment programme would improve specialist education. Fourteen consultants mentioned problems related to the implementation of the programme or thought that the programme or part of it was too ambitious. There were a number of comments about the wording of items. A few

consultants mentioned the absence of specific skills, such as laryngeal mask and pain treatment.

Discussion

The consultants' opinion of the new assessment programme is important for various reasons. External opinion about the representativeness and the content of an assessment programme is essential before implementation of any new assessment programme.[7, 8] Second, the introduction of formal clinical assessment is new and controversial in Denmark, where the prevailing assessment method in residency training has been formative with no end-of-training examination.[9]

There was strong agreement regarding the representativeness of the tests included in the assessment programme. Any assessment programme will only be testing a subset of all the expected skills and competences. Surprisingly few consultants mentioned skills or competencies that they missed in the programme. The questionnaire did not include any specific test of how careful the consultants evaluated the different proposals. The number of comments about the wording of items and spelling errors indicate that the consultants actually were very careful. The fact that the assessment programme is practice-based and will be used by consultants nationwide also gives us reason to believe that they were rather careful.

Strong agreement was observed regarding the content of most of the tests. In two tests the lower quartile was 3. One was the test *managing operations list on call*. The other test reflecting this competency, *teamwork and interpersonal skills*, showed strong agreement, which indicates that the perceived importance of this competency is not the problem. It is more likely that the variance reflects differences in work organisation in individual departments, with differences in the amount of responsibility given to first-year residents while on call. The other test, *'Evidence-based-medicine exercise'*, where the lower quartile was 3, had a tendency to be rated as being of lower importance by the consultants from general hospitals than by those from university hospitals. This is not surprising as research activities are more prevailing among consultants at university hospitals, and hence this competency is likely to be more familiar.[10] Consultants at general hospitals may have doubts concerning the presence in their departments of adequately qualified supervisors and assessors for these assignments. Although this concern was not expressed in the respondents' comments, there may be a need for training of trainers.

There was surprisingly positive attitude towards the programme. Denmark has previously resisted implementing the European Board Ex-

amination in Anaesthesiology or a similar examination despite their use in several European countries.[11, 12] The reason for the high degree of acceptance of our programme may be due to it being perceived as an appealing alternative to traditional final examination formats.

This study is limited to consultants' opinions on the outline of the assessment programme and not on actual experience with the programme. Further studies are needed to assess the feasibility of the programme and its acceptance in practice by both consultants and residents.

Conclusion

Judging from the consultants' opinions, the assessment programme for the first-year residency in anaesthesiology appears to be appropriate regarding the range of competencies assessed, the appropriateness of the tests for making pass/fail decisions and regarding the relevance and sufficiency of the content of each of the 18 tests used for pass/fail decisions.

Acknowledgements: The working group that compiled the curriculum and the assessment programme for first-year residency in anaesthesiology included: Andersen N, Berlac PA, Bested K, Callesen T, Christensen P, Jensen E, Jensen JW, Lemholt K, Lund J, Malling B, Mandøe H, Nørregaard O, Pedersen BD, Petersen JA, Ravlo O, Ravn L, Ringsted C, Skjelsager K, Sprehn M, Østergaard D. We want to thank the consultants in anaesthesiology for answering the questionnaire and their constructive comments.

References

1. Friedman Ben-David M. The role of assessment in expanding professional horizons. Med Teach 2000; 22: 472-7.
2. CanMEDS 2000: Extract from the CanMEDS 2000 Project Societal Needs Working Group Report. Med Teach 2000; 22: 549-54.
3. Van der Vleuten CPM. The assessment of professional competence: developments, research and practical implications. Adv Health Sci Educ 1996; 1: 41-67.
4. Shepard LA. The role of assessment in a learning culture. Educ Research 2000; 29: 4-14.
5. The National Board of Health. Guidelines for Curriculum of Specialist Education. Copenhagen: The National Board of Health, 2001.
6. Ringsted C, Østergaard D. The education book for first year residency in anaesthesiology. [Uddannelsesbog. Introduktionsuddannelsen i anæstesiologi, H:S Postgraduate Medicinske Institut] (www.dasaim.dk and www.hspmi.dk). CHC Postgraduate Medical Institute, Copenhagen, 2001.
7. Ebel RL, Frisbie DA. Essentials of educational measurement. In: Ebel RL, Frisbie DA, eds. 5th edn. Englewood Cliffs, New Jersey: Prentice Hall, 1991: 102-5.
8. Fowell SL, Southgate LJ, Bligh JG. Evaluating assessment: the missing link? Med Educ 1999; 33: 276-81.
9. Karle H, Nystrup J. Comprehensive evaluation of specialist training: an alternative to Board examinations in Europe? Med Educ 1995; 29: 308-16.
10. Nielsen LB, Jensen TUS, Skovgaard LT, Viby-Mogensen J. Research in Danish departments of anaesthesiology at the turn of the century: A biblometric analysis. Ugeskr Laeger 2001; 163: 6121-7.
11. Wiklund L. Education of specialists in anaesthesia and intensive care in Europe - A Swedish perspective. Ann Acad Med Singapore 1994; 23: 623-9.
12. Zorab JS. The European Diploma in Anaesthesiology and Intensive Care of the European Academy of Anaesthesiology. Acta Anaesthesiol Scand 1995; 39: 579-81.

Appendix 1. Assessment programme for first-year residents in anaesthesiology

	Competence	When?	How?	Summative or formative *
1.	Managing airways, intubation	• Before on call in anaesthesia	• Observation by checklist + questions	• Summative
2.	Testing the anaesthesia machine	• Before on call in anaesthesia	• Observation by checklist + questions	• Summative
3.	Principles of general anaesthesia	• Before on call in anaesthesia	• Observation by checklist + questions	• Summative
4.	Principles of general anaesthesia to emergency patient	• Before on call in anaesthesia	• Observation by checklist + questions	• Summative
5.	Preoperative assessment of patient	• Before on call in anaesthesia	• Observation by checklist + questions	• Summative
6.	Advanced CPR	• Before on call in anaesthesia	• Observation on mannequin, checklist + questions	• Summative
7.	Management plan for fluids and nutrition	• Before on call in intensive care	• Audit on three patient plans + questions	• Summative
8.	Managing simple ventilator	• Before on call in intensive care	• Observation by checklist + questions	• Summative
9.	Managing simple intensive care patient	• Before on call in intensive care	• Observation by checklist + questions	• Summative
10.	Spinal anaesthesia	• Whenever suitable	• Observation by checklist + questions	• Summative
11.	Epidural anaesthesia	• Whenever suitable	• Observation by checklist + questions	• Summative
12.	Central venous catheter	• Whenever suitable	• Observation of checklist + questions	• Summative
13.	Registration of experience in anaesthesia and procedures	• At 1st, 3rd, 6th, 9th, and 11th month	• Continuous registration, discussion with tutor	• Formative

	Competence	When?	How?	Summative or formative *
14.	Registration of quality of spinal and epidural anaesthesia, central venous and arterial line	• At 1st, 3rd, 6th, 9th, and 11th month	• Continuous registration, discussion with tutor	• Formative
15	Patient communication skills	• After 6 months	• Patient survey, administered by the resident, results assessed by tutor	• Summative
16.	Team work and interpersonal skills	• After 6 months	• Report from two supervisors according to checklist, discussion with tutor	• Summative
17	Managing operating list on call	• After 6 months	• Assessment by senior colleague according to checklist and questions to the team.	• Summative
18.	Anaesthesia and complicating medical diseases	• Whenever suitable	• Written assignment, assessment by supervisor	• Summative
19.	Patient management plan, a critical appraisal of practice	• Whenever suitable	• Written assignment, assessment by supervisor	• Summative
20.	Evidence-based medicine exercise	• Whenever suitable	• Written assignment, assessment by supervisor	• Summative
21.	Learning portfolio	• At 1st, 3rd, 6th, 9th, and 11th month	• Discussion with tutor	• Formative

* Summative means used for pass/fail decisions. Formative means used for monitoring the resident's progress and providing feedback to the resident to guide further learning.

CHAPTER 3

Feasibility of scoring forms

Published in Medical Teacher 2003; 25: 654-8. http://tandf.co.uk/journals

A feasibility study comparing checklists and global rating forms to assess resident performance in clinical skills.

C Ringsted, D Østergaard, L Ravn, JA Petersen, PA Berlac, CPM van der Vleuten

Abstract

This study evaluated the feasibility of two different scoring forms for assessing the clinical performance of residents in anaesthesiology. One of the forms had a checklist format including task specific items and the other was a global rating form with general dimensions of competence including 'clinical skills', 'communication skills', and 'knowledge'. Thirty-two clinicians representing 25 (83%) of the 30 training hospitals in the country participated in the study. The clinicians were randomised into two groups each of which used one of the scoring formats to assess a resident's performance in four simulated clinical scenarios on videotape. Clinicians' opinions about the appropriateness of the scoring forms were rated on a scale of 1-5. The checklist format was rated significantly higher compared with the global rating form (mean 4.6 SD 0.5 vs. mean 3.5 SD 1.4, $p<0.001$). The inter-rater agreement regarding pass/fail decisions was poor irrespective of the scoring form used. This was explained by clinicians' leniency as assessors rather than by lack of vigilance in the observations or disagreements on standards for good performance.

Introduction

The introduction of in-training assessment using clinicians to assess the performance of residents is a challenge. The assessment method must be easy to both use and understand and must also provide reliable results. For technical skills task specific checklists have been shown to be reliable measurement tools in standardised settings.[1] For complex tasks in clinical settings checklists tend to be too elaborate to cover all subsets of the task. For example Sivarajan et al.[2] used 61 items in a checklist for epidural anaesthesia. This might not be a feasible instrument in a clinical setting and it is a lot of work to design if several different tasks are to be observed. One alternative to the use of checklists could be a rating form with general dimensions of competence such as those used for clinical observation of residents' patient encounters in internal medicine.[3] The advantage of such a form is that it adapts to a broad range of clinical situations.

The aim of this study was to evaluate the feasibility of two structured scoring formats, that is, checklists including task specific items versus global rating forms with general dimensions of competence including clinical skills, communication skills and knowledge. We wanted to study whether checklists were superior to global forms for observation of more technically oriented clinical tasks with respect to acceptability and inter-rater agreement on pass/fail decisions.

Context of the study

A working group under the Danish Society of Anaesthesiology and Intensive Care Medicine had developed a new nationwide in-training-assessment programme for first year residency of anaesthesiology.[4] The programme included twenty-one individual tests on various clinical tasks and the tests were spread out over the entire first year of residency. The programme was to be administered at the individual departments using clinicians as assessors for the individual tests. The working group, before introduction of the programme, established a consensus standard for pass/fail decisions implying that all specific tests must be passed in order for the resident to pass the entire programme. For each of the individual test the criteria for passing a resident was mastery meaning that all elements of the performance must be correct. The standard and the criteria were confirmed in a survey among consultants in anaesthesiology.[5] The design of the program includes provisions so that if a resident fails a test, proper remediation is instituted and the assessment repeated in due time. The final report on each resident should only include the tests passed successfully. Hence study of test reliability using residents' actual test results was not feasible. We therefore designed an experimental study where clinicians score simulated videotaped scenarios rather than actual test situations.

Methods

Four different test scenarios were simulated and videotaped. Each scenario demonstrated a resident performing a clinical task and a supervisor observing and questioning the resident. The scenarios included insertion of an epidural catheter for anaesthesia, a preoperative anaesthesia consultation, an emergency induction of general anaesthesia, and a round on a patient in the intensive care unit. A group of expert consultants were asked to define two mistakes in each of the first three scenarios that included one 'sloppiness' and one 'serious error' in each scenario, which to the experts' opinion should result in a clear fail of the resident. In each of the first three scenarios the simulated resident was instructed to make these two specific mistakes. For example, in the epidural scenario this was being sloppy with the skin disinfection and withdrawal of the epidural catheter while still inside the needle. In the preoperative consultation scenario the mistakes were not questioning about status of teeth and not performing the relevant physical examination of the patient. In the emergency induction scenario the resident did not check appropriately for patient identification and second, did not perform crash induction cor-

rectly. The intensive care scenario did not include any deliberate mistakes.

Two different scoring forms were elaborated for assessing the four different clinical tasks. One was a task specific checklist including from nine to 14 items with a dichotomous scoring of each item, *yes/no*, and an overall dichotomous scoring of pass/fail. All checklists included items on communication skills, technical performance and some theoretical questions related to the task. An example of a checklist is shown in Appendix 1. The second scoring form was a global form that included three general aspects of competence to be assessed: communication skills, clinical skills and cognitive knowledge. The form contained for each of these categories a five point scoring, 1=clear fail, 2=borderline fail, 3=borderline pass, 4=clear pass, 5=excellent. This form included an overall dichotomous scoring of pass/fail. The global form is shown in Appendix 2.

A letter was sent to the anaesthetic department at each of the 30 training hospitals across the country inviting the educational responsible consultant at the department and/or one of the clinical supervisors to participate in a two-hour study about assessment of residents' performance. The participants were instructed to use their professional expertise and judgement in evaluation of the resident's performance depicted in each of the four video taped scenarios. The assessors were also instructed to make an overall decision regarding pass or fail of the resident according to the assessment protocol. According to the protocol the criteria for passing a residents' performance on each scenario was mastery meaning that all elements of the performance must be correct.

The participants were randomised into two groups. Group one used checklists specific to the tasks in each of the four scenarios. Group two used identical global scoring forms for each scenario. For each scenario the participants were asked to indicate the appropriateness of the scoring form they had used for assessment of resident performance. Appropriateness was indicated on a scale 1 to 5 (1=not at all appropriate, 5=very appropriate). Also on each form the participants were invited to give written comments. The participants were not asked specifically to indicate observed mistakes. Each video lasted from ten to fifteen minutes. Time was given between each of the four scenarios to fill in comments.

Statistical analysis

For each examiner and scenario an arbitrary index of the resident's mean performance score was calculated as a percentage of correct items on the checklists and as a percentage of the total rating of the three dimensions of competence on the global forms. Comparisons between groups were made using Kruscal-Wallis test for comparing data of mean performance score and appropriateness of scoring form. The chi-squared test was used

for comparison of categorical data, p-values below 0.05 were considered statistically significant.

Results

Fourteen educational responsible consultants and eighteen clinical supervisors representing 25 (83%) of the 30 teaching hospitals accepted the invitation. There was an equal representation of consultants and supervisors and an equal geographical representation in the two groups.

The appropriateness of the checklists was rated statistically significantly higher than the global format in each of the four scenarios (Table 1), (overall mean 4.6 SD 0.5 vs. mean 3.5 SD 1.4, $p<0.001$). Several participants commented that a checklist is of great help when assessing a resident.

The pass/fail decisions on each of the four scenarios are shown for the two groups of assessors in Table 1. In all of the first three scenarios where the resident makes two deliberate mistakes there was no statistical significant difference in number of clinicians passing or failing the resident in either of the two groups. In the fourth scenario where the resident was not instructed to make mistakes all assessors in both groups passed the resident. There were no significant differences between groups in the calculated mean performance score of any of the four scenarios.

The majority of assessors using checklists registered one or both of the mistakes in the first three scenarios, 94% (15/16), 100% (16/16), and 81% (13/16) and several assessors using the global format indicated the mistakes in the commentaries, 69% (11/16), 38% (6/16), and 25% (4/14). Several assessors in both groups passed the resident despite noticing the mistakes as indicated either in the scoring of the items of the checklist or in the commentaries. On 10 occasions the assessors indicated that they passed the resident despite mistakes, but that they would give some corrections to the resident if this were a real life situation.

Table 1. Thirty-two clinicians' assessment of four simulated resident performances.

Scenario	Group 1 Checklist form	Group 2 General format	Total
Epidural anaesthesia			
Number of clinicians passing the resident	6	7	13
Number of clinicians failing the resident	10	9	19
Total	16	16	32
Calculated mean performance score (SD), percentage	72.2 (14.0)	67.9 (6.2)	
Appropriateness of scoring form	4.6 (0.5)	3.6 (1.4)*	
Pre-operative consultation			
Number of clinicians passing the resident	8	10	18
Number of clinicians failing the resident	8	6	14
Total	16	16	32
Calculated mean performance score (SD), percentage	72.4 (17.7)	71.2 (11.2)	
Appropriateness of scoring form	4.5 (0.6)	3.3 (1.3)*	
General anaesthesia			
Number of clinicians passing the resident	10	10	20
Number of clinicians failing the resident	6	4	10
Total	16	14§	30
Calculated mean performance score (SD), percentage	80.0 (11.0)	72.1 (12.8)	
Appropriateness of scoring form	4.6 (0.5)	3.4 (1.4)*	
Intensive care patient			
Number of clinicians passing the resident	16	16	32
Number of clinicians failing the resident	0	0	0
Total	16	16	32
Calculated mean performance score (SD), percentage	97.2 (6.4)	82.8 (8.9)	
Appropriateness of scoring form	4.9 (0.4)	3.2 (1.2)*	

Notes: Group one used a checklist with task specific items and group two used a form with general dimensions of competence for the assessment. The table shows the number of assessors in each group and in total that passed or failed the resident, the calculated mean performance score (SD), and the clinicians' mean rating of appropriateness. * Checklist form rated significantly higher than general format, $p<0.01$. § Two assessors in the general format group did not make an overall decision of pass/fail in scenario three.

Discussion

This study demonstrates that the clinicians found checklists more appropriate than global rating forms with categories of competence when scoring residents' performance on complex clinical tasks. This finding was consistent across all four clinical scenarios, although checklists could have been expected to be more appropriate for technically oriented tasks and general scoring forms more appropriate for patient encounters. In OSCE-examinations and standardised assessment using trained expert examiners several studies have shown that global rating might be a better alternative to checklists in terms of reproducibility and efficiency in test development and administration.[1, 6-8] However, for in-training assessment, checklists might be a better choice because they can help make the program's learning objectives more explicit to both residents and supervisors.

The study showed a poor agreement among clinicians regarding pass/fail decisions irrespective of scoring form used. This poor agreement was probably not a problem of accuracy in observation of the resident's performance. The fact that the mean score in any of the scenarios was not significantly different between the two groups and that a number of clinicians in both groups noticed the deliberate mistakes even among those who passed the resident, indicates that the vigilance was equally high in both groups. Furthermore that all clinicians passed the resident in the last scenario without deliberate mistakes supports the validity of the videos. The contradiction of passing a resident performing incorrectly was probably not a question of underestimating the mistakes as several assessors emphasized the severity of the mistakes in their comments and indicated that they would address this in real life situations and instruct the resident to correct the performance. Thus irrespective of scoring form this study indicates a problem with assessors' consistency and compliance with assessment protocol. The study of Wilkinson *et al.*[9] who found that station construction and mark sheets contributed 10.1% and the examiner 89.9% to the variation in inter-rater reliability. This is in accordance with our findings and similar results of disagreement on pass/fail decisions in other studies using clinicians as assessors.[3, 10, 11] Noel *et al.*[11] demonstrated that the accuracy in evaluating residents' clinical performance by direct observation could be enhanced if assessors used a structured form, but still they found profound disagreement regarding overall pass/fail decisions. Kroboth *et al.*[3] performed an extensive reliability study on scoring forms with categories of competence and also found low inter-rater reliability.

Our study was performed at the initial introduction of formal assessment in postgraduate education in the country and unfamiliarity with the

concept might in part explain the results and point to the need to train the clinicians as assessors. However, Wilkinson *et al.*[9] demonstrated that examiner experience as assessors was not associated with inter-rater reliability in OSCEs. It has previously been shown that training assessors is ineffective for inconsistent assessors and that simply removing these from the analysis of the results on OSCE exams will increase reliability.[12]

Due to problems of reliability in performance-based assessment it is recommended to have several different assessors and several observations with a least amount of structure in the test formats.[13] Using these principles reliable instruments for assessing residents' performance in a clinical context has been elaborated.[14-16] However, these studies probably also have the problem of lenient assessors. In the studies on the Mini-Clinical evaluation exercise (mCEX) for internal medicine residents where several different attending physicians observe the residents on several different patient encounters using scoring forms with general aspects of competence, only small differences in ratings across examiners and settings were found.[14, 15] Especially in the study by Durning *et al.*[14] the mean score on seven consecutive mCEX during the residents' first year were not significantly different and did not indicate progress in performance over the year as expected. As opposed to these two studies Yudkowsky[16] used an approach that demonstrated an incremental growth in residents' competence. Two things probably contributed to this. First, they used a developmentally-oriented rating scale emphasising the progress expected in terms of the amount of supervision required to perform the task. Second, the decision of satisfaction with performance was being left to an education committee and not to the clinician.

Although our assessment programme for first year residents in anaesthesiology is in accordance with the principles using multiple observations and assessors[4] the results of this study point to the need for further study on how an in-training assessment programme works in clinical practice.[13, 17]

Conclusion

Clinicians found checklists with task specific items more appropriate than rating forms with general dimensions of competence for the purpose of scoring residents' clinical performance. Consistency of pass/fail decisions was equally poor for checklists compared to global scoring forms. This was explained by clinicians' leniency as assessors rather than by lack of vigilance in the observations or disagreements on standards for good performance

Acknowledgements: This study was supported in part by a grant from the Ministry of Health, Denmark. The authors wish to thank the clinicians for taking time to participate in this study.

References

1. Morgan PJ, Cleave-Hoss D, Guest CB. A comparison of global ratings and checklist scores from an undergraduate assessment using an anesthesia simulator. Acad Med 2001; 76: 1053-5.
2. Sivarajan M, Miller E, Hardy C, Herr G, Liu P, Willenkin R, Cullen B. Objective evaluation of clinical performance and correlation with knowledge. Anesth Analg 1984; 63: 603-7.
3. Kroboth FJ, Hanusa BH, Parker S, Coulehan JL, Kapoor WN, Brown FH, Karpf M, Levey GS. The inter-rater reliability and internal consistency of a clinical evaluation exercise. J Gen Int Med 1992; 7: 174-9.
4. Ringsted C, Østergaard D, Scherpbier A. Embracing the new paradigm of assessment in residency training: an assessment programme for first-year residency training in anaesthesiology. Med Teach 2003; 25: 54-62.
5. Ringsted C, Østergaard D, Scherpbier A. Consultants' opinion on a new practice based assessment programme for first year residents in anaesthesiology. Acta Anaesthesiol Scand 2002; 46: 1119-23.
6. Govaerts MJB, Van der Vleuten CPM, Schuwirth LWT. Optimising the reproducibility of a performance-based assessment test in midwifery education. Adv in Health Sci Educ 2002; 7: 133-45.
7. Hodges B, Regehr G, McNaughton N, Tiberius R, Hanson M. OSCE checklists do not capture increasing levels of expertise. Acad Med 1999; 74: 1129-34.
8. Regehr G, MacRae H, Reznick RK, Szalay D. Comparing the psychometric properties of checklists and global rating scales for assessing performance on an OSCE-format examination. Acad Med 1998; 73: 993-7.
9. Wilkinson TJ, Frampton M, Thompson-Fawcett M, Egan T. Objectivity in objective structured clinical examinations: Checklists are no substitute for examiner commitment. Acad Med 2003; 78: 219-23.
10. Elliot DL, Hickam DH. Evaluation of physical examination skills. JAMA 1987; 258: 3405-8.
11. Noel GL, Herbers JE, Caplow MP, Cooper GS, Pangaro JN, Harvey J. How well do internal medicine faculty members evaluate the clinical skills of residents? Ann Int Med 1992;117: 7 57-65.
12. Newble DI, Hoare J, Sheldrake PF. The selection and training of examiners for clinical examinations. Med Educ 1980; 14: 345-9.
13. Van der Vleuten CPM. The assessment of professional competence: Developments, research and practical implications. Adv Health Sci Educ 1996; 1: 41-67.
14. Durning SJ, Cation LJ, Markert RJ Pangaro LN. Assessing the reliability and validity of the Mini-clinical evaluation exercise for internal medicine residency training. Acad Med 2002; 77: 900-4.
15. Norcini JJ, Blank LL, Arnold GK, Kimball HR. Examiner differences in the Mini-Cex. Adv Health Sci Educ 1997; 2: 27-33.
16. Yudkowsky R. Can resident evaluations demonstrate increases in residents' skills over time? Acad Med 1999; 74: S108-10.
17. Fowell SL, Southgate LJ, Bligh JG. Evaluating assessment: the missing link? Med Educ 1999; 33: 276-81.

Appendix 1. Example of checklist scoring form used for assessing a resident's performance.

Insertion of epidural catheter.	YES	NO
Informs and guides patient appropriately regarding the procedure and positioning	☐	☐
Correct positioning of patient, sitting or prone position, with assistance from staff	☐	☐
Correct preparation of utensils and check of drugs and instruments	☐	☐
Proper identification of insertion level appropriate for the planned operation	☐	☐
Correct disinfection and sterile draping of skin	☐	☐
Discuss choice of local anaesthetic for skin and choice of median or para-median approach	☐	☐
Correct insertion of needle	☐	☐
Correct identification of epidural space – loss of resistance	☐	☐
Correct insertion of epidural catheter	☐	☐
Test of catheter position – can explain the reason and procedure for testing	☐	☐
Discuss and defend choice and dose of anaesthetic for the epidural block	☐	☐
Determines block level -with indication of dermatomes	☐	☐
Describes indications and contra-indications for epidural anaesthesia	☐	☐
Describes at least three important complications, describe how to prevent and treat these	☐	☐
I will pass the resident regarding this competency	☐	☐

Appendix 2. Example of global scoring form with general dimensions of competence used for assessing a resident's performance.

Insertion of epidural catheter

The resident's performance regarding patient communication is:

Clear fail	Borderline fail	Borderline pass	Clear pass	Excellent
☐	☐	☐	☐	☐

The resident's performance regarding clinical and technical skills is:

Clear fail	Borderline fail	Borderline pass	Clear pass	Excellent
☐	☐	☐	☐	☐

The resident's performance regarding knowledge is:

Clear fail	Borderline fail	Borderline pass	Clear pass	Excellent
☐	☐	☐	☐	☐

	YES	NO
I will pass the resident regarding this competency	☐	☐

CHAPTER 4

Implementation of ITA

Published in Acta Anaesthesiologica Scandinavica 2003; 47: 1196-1203

Implementation of a formal in-training-assessment programme in anaesthesiology and preliminary results of acceptability.

C Ringsted, D Østergaard, CPM van der Vleuten

Abstract

Background: A new reform on postgraduate education in Denmark requires formal in-training assessment in all specialties. The aim of this study was to survey the implementation and acceptability of the first example of a nation-wide in-training-assessment programme for first year trainees in anaesthesiology developed by a working group under the Danish Society of Anaesthesiology and Intensive Care Medicine.
Methods: A questionnaire about the implementation of the programme in practice and the characteristics of trainees was sent to the educational responsible consultant (ERC) in each of the 26 anaesthetic departments in the country with first year trainees in anaesthesiology. Standard evaluations of the assessment programme were regularly collected from trainees.
Results: Twenty-five (96%) departments returned the questionnaire. In total the departments reported on 100 trainees and 83 of these had been enrolled in the programme. Thirteen departments reported in total on 27 trainees who had completed their first year training and these departments had applied median 21 (range 17-21) of the 21 tests included in the entire programme. Time constraints and resistance among senior clinicians were the most frequently cited barriers to implementation. Evaluations from trainees showed a general positive attitude towards most of the programme. They especially praised the programme's effect on structuring training and having a positive effect on learning.
Conclusion: The in-training-assessment programme has been widely implemented across the country. The majority of the programme was acceptable to trainees and had a positive effect on structuring training and on fostering learning.

Introduction

Since 2000 a new reform in postgraduate education in Denmark has been under way.[1, 2] The new reform reflects the general paradigm shift towards concepts of outcome based education and comprehensive performance-based assessment as an on-going process rather than end-of-training assessment.[3-6] Apart from medical expertise this includes interpersonal, organisational, and academic competence and professionalism.[2-5] The need for more structured training and assessment in all these aspects has repeatedly been demonstrated.[7-10] Introducing a formal in-training assessment is an extra challenge in Denmark as until now there has been no formal clinical assessment of trainees and no specialist exams in any specialties.[11, 12]

This study deals with the implementation of the first example in Denmark of a nation-wide curriculum and in-training assessment programme complying with the new guidelines and the paradigm of practice-based assessment as an on-going process.[13, 14] The programme was for first year trainees in anaesthesiology and was developed by a working group under the Danish Society of Anaesthesiology and Intensive Care Medicine (DASAIM).[13] The working group of 20 people included members of the DASAIM's board of education, consultants from various parts of the country, and trainees. The programme was developed through several seminars and literature studies and all decisions were made by consensus in the working group. A draft was distributed to core faculty for review and a nation-wide survey on consultants' opinion showed high agreement regarding the scope, content, and importance of the elements of the programme.[14]

The aim of this study was to survey the implementation of the programme across the country and especially to identify any barriers to implementation. Also we wanted to obtain a first impression of the programme's acceptability to trainees.

Methods

The programme

The assessment programme includes 21 individual elements spread out through the first year of training and tailored to practise and the trainee's professional development (Table 1).[13] Eighteen of the elements are summative tests to which are associated pass/fail decisions. These include 12 observations of trainee performance on clinical tasks for which the supervisor uses a competence-card composed of a checklist of task specific items and some questions on related theory. One element is one a survey on the trainee's communication skills, distributed to 25 patients by the trainee during preoperative consultations. Two elements are structured reports from attending anaesthetists, one in daytime and one on-call, for assessment of the trainee's organisational and interpersonal behaviour. Three instruments are reflective written assignments on patient cases, one being an exercise of evidence-based medicine.

Table 1. Specification of 21 elements included in the assessment programme for first year trainees in anaesthesiology.

Clinical skills 1. Managing airways 2. Testing anaesthesia machine 3. General anaesthesia 4. Emergency induction of anaesthesia 5. Preoperative consultation 6. Advanced resuscitation 7. Plan for fluid and nutrition 8. Function of standard respirator 9. Managing intensive care patient 10. Spinal anaesthesia 11. Epidural anaesthesia 12. Central venous catheter	These tests are performed by having a supervisor observe the trainee in practice and score the performance by use of a competency-card with task specific items regarding performance and some questions about theory. Each test includes an overall decision regarding pass/fail.
Experience 13. Recording of experience 14. Cusum-scoring	Recording of experience includes various kinds of patients and procedures. An electronic database was available on the Internet. Cusum-scoring is registration of success or failure rate each time a procedure is performed. Four procedures were planned for Cusum-scoring, spinal anaesthesia, epidural anaesthesia, arterial catheter, and central venous catheter. Success and failure rate is specified taking into account the 'normal' failure rate for expert anaesthetists. Electronic database available.
Communication 15. Survey on patient communication	The trainee hands out a questionnaire to 25 patients regarding the preoperative consultation and elaborates a critical report on the results. The supervisor assesses the report and makes a pass/fail decision.
Managerial and collaborative skills 16. Organisation, collaboration, professionalism on call 17. Organisation, collaboration, professionalism in OR	Senior clinicians, that have worked closely with the trainee, rate the trainee on a structured form, one consultant's rating in test no 16. and two in test no 17. The trainees' clinical supervisor makes an overall decision on behalf of the ratings regarding pass/fail on each of two tests.
Written assignments 18. Choice of anaesthetics 19. Reflection on case 20. Evidence based medicine	Written assignments are scored according to specified criteria and an overall pass/fail decision is made for each test.
Self-directed learning 21. Learning portfolio	The trainee identifies individual learning objectives, means to reach the goals and reports on the learning. The supervisor monitors the progress of the trainee's learning and gives an overall mark on the professional development according to specified criteria on a checklist.

For each of these 18 summative elements the supervisor makes an overall decision regarding pass or fail. In case the trainee fails, proper remediation should be instituted and the test repeated in due time. All 18 elements must be successfully passed in order to have the first year training approved. The last three instruments had a formative purpose aiming at guiding training and learning. Two of these are continuous recording of experience in logs and the third is a learning portfolio. One of the logs is recording of experience with patients and procedures. The other one specifically monitors the quality of the performance of four technical procedures, by use of the Cusum technique.[15] The principle in the Cusum technique is that every time the procedure is performed, the trainee records success or failure by a specific scoring system, which takes the failure rate of experienced anaesthetists into account.

General implementation strategy

The curriculum and the assessment programme were issued in a booklet and sent to all anaesthetists in the country. A questionnaire accompanied the booklet to all consultants asking for their opinion about the content of the programme in general and of the content of each competence-card.[14] A handbook on assessment protocols was elaborated and sent to all departments. The working group ran several information meetings across the country inviting all departments to participate. Barriers to implementation of the written assignments in the programme were foreseen [13, 14] and hence special workshops for supervisors were conducted on how to manage the written assignment and principles of evidence based medicine. Finally the programme was introduced to the trainees at the regional mandatory seminars that accompanied the clinical training programme and special emphasis was put on how to perform assignments on evidence-based medicine.

Survey on the implementation in practice

This study was performed 1.5 years after the introduction of the programme. A questionnaire about the implementation of the programme in the departments was sent to the educational responsible consultant (ERC) in each of the 26 anaesthetic departments with first year trainees. Departments were asked to report the number of trainees and to indicate dates of employment for each trainee, and to what extent they had been enrolled in the assessment programme: completely, partly or not at all. Further each department was asked to tick off, which of the 21 elements in the assessment programme they had applied. The departments were asked to indicate reasons for not applying specific elements of the programme or the entire assessment programme if that was the case. Finally, departments were invited to give comments of any kind.

Trainees' evaluation
An evaluation form regarding trainees' opinion about the assessment programme was included in the booklet about the programme. The form included ratings of specific aspects of the programme on a scale of 1-9, 1=not at all/very bad and 9=very much/very good. The form also included open questions regarding positive and negative aspects of the programme. The trainees were asked to fill in the evaluation forms after termination of the entire assessment programme and send it to the authors. Trainees' identity were not known to the authors or included in any common database and hence personal reminder procedures were not possible. A general reminder was sent twice to the departments.

Results

Implementation
The questionnaire was returned by 25 (96%) of the 26 departments. Four departments reported not having any trainees. In total, 21 departments reported on 100 trainees. Twenty-seven trainees had completed the first year training and all of these had been enrolled in the assessment programme. Fifty-one trainees had only had part of their training during the survey period and 49 of these had been enrolled in the programme. Finally there were 22 trainees who had only had short-term employments either as a vacancy or as a supplementary training for another specialty, typically for 3-6 months. Seven trainees in this group had been enrolled in parts of the assessment programme. In total, 83 trainees had been enrolled in the programme.

One of the 21 departments with only one recently started resident did not report on application of the elements of the assessment programme. Twenty departments reported on their application of the 21 elements of the entire assessment programme (Table 2). Thirteen departments reported on trainees who had completed their training and these departments had applied median 21 (range 17-21) of the individual 21 elements included in the entire programme. Seven departments reported on only having trainees who had had part of the training and these departments had applied median 12 (range 6-20) of 21 elements. The twelve clinical skills tests were the most widely implemented elements of the programme. The two least implemented elements of the programme were the test on communications skills and the learning portfolio.

Table 2. Application of the 21 individual elements included in the assessment programme for first year trainees in anaesthesiology in departments with trainees who had completed the programme and departments only having trainees who had had part of their training. Questionnaire returned from 25 of 26 departments in Denmark. [5]

	Number of departments reporting on applying the individual elements of the programme	
The 21 elements included in the assessment programme	Departments with trainees who had completed their training (N=13)	Departments only having trainees who had had part of their training. (N=7)
Clinical skills		
1. Managing airways	13	7
2. Testing anaesthesia machine	13	7
3. General anaesthesia	13	7
4. Emergency induction of anaesthesia	13	4
5. Preoperative consultation	13	5
6. Advanced resuscitation	13	6
7. Plan for fluid and nutrition	13	4
8. Function of standard respirator	13	2
9. Managing intensive care patient	13	3
10. Spinal anaesthesia	13	7
11. Epidural anaesthesia	13	6
12. Central venous catheter	13	4
Experience		
13. Continuous registration of experience	12	2
14. Cusum-scoring	12	2
Communication		
15. Survey on patient communication	11	3
Managing and collaborative skills		
16. Organisation/collaboration on call	13	3
17. Organisation/collaboration in OR	12	3
Academic competence		
18. Written assignment, choice of anaesthetic	12	3
19. Written assignment, reflection on case	12	3
20. Written assignment, evidence based medicine	12	2
Self-directed learning		
21. Learning portfolio	10	1

[5]Four departments had no trainees in the period. One department with one recently started resident did not report on application of the individual elements of the assessment programme.

Departments' comments in the questionnaires

Six departments praised the programme on several aspects: for making goals and objectives clear to both trainees and trainers; for being of help in structuring the training; for emphasising responsibility of both the trainee and the supervisor; and for documentation of competence and monitoring progress. One department used this documentation to reward the trainees with greater responsibilities in practice. The departments commented that the trainees were enthusiastic about the programme. One department reported on having been awarded the county's annual 'Education reward' with special praise to the programme.

One of the barriers to implementation was difficulty in motivating the senior clinicians to engage in the assessment of trainees and five departments mentioned that. Also, five departments commented on the conflict of service demands and time needed for the assessments. Specific reasons for not applying the survey of trainee's communication skills included not finding it worthwhile considering the efforts. One department found the two tests on organisational skills/ collaboration to be a duplicate and had gathered them in one. Several departments reported that senior clinicians found the written assignments a big challenge and one department used the principles but with oral presentations and discussions instead of writing. Recording of experience was problematic due to technical problems with the electronic database. Some departments did not use the learning portfolio as intended and one department commented that the assessment programme in itself was some sort of a portfolio.

Trainees' evaluations

Fifteen trainees from eight different departments had completed and returned the evaluation form (Table 3). Results showed that, in general, the booklet and the assessment programme were rated high (median at or above 7 on a scale of 1-9). The recording of experience, the survey on communication skills, and the learning portfolio, were less popular instruments compared to the tests on clinical skills, the test on organisational and collaborative skills, and the written assignments.

Table 3. Trainees' evaluation of an in-training-assessment programme for first year training in anaesthesiology. Rating of items on a scale 1-9, 1=not at all/very bad and 9=very much/very good. Results are given as median (range).

	Residents (N=15)
General opinion about the booklet	
Did the booklet on the curriculum and the assessment programme in general include sufficient information?	8 (6-9)
Was the booklet easy to understand and well structured?	7 (5-9)
Were the booklet and the assessment useful for the appraisal meetings with the supervisor?	7 (3-9)
General opinion about the assessment programme	
Did you use the competency-cards in guiding your learning?	8 (4-9)
In general has the assessment programme been useful?	7 (5-9)
Has the assessment contributed to your self-confidence and insight regarding your competence?	7 (1-9)
Specific opinion about the individual tests	
Were the tests on clinical skills (no 1-12) an acceptable way to be assessed?	8 (4-9)
Was the registration of experience (no 13) useful considering the effort put into it?	5 (1-9)*
Was Cusum-scoring (no. 14) useful considering the effort put into it?	8 (1-9)
Was the survey on communication skills (no 15) an efficient learning experience?	5 (2-9)*
Were the tests on organisational skills/collaboration (no 16-17) an acceptable way to be assessed?	8 (4-9)
Were the written reflective assignments (no 18-20) a worthwhile learning experience considering the effort?	7 (4-9)
Was the learning portfolio (no 21) useful in structuring the training and learning?	5 (1-9)*

* Item scored statistically significant lower compared to 'clinical skills test', 'tests on organisational skills/collaboration, and the 'written assignments', Wilcoxon signed ranks test, $p<0.05$.

Trainees' responses to the open-ended questions (Table 4) showed that the positive aspects of the programme were making goals and objectives very clear and broadening the aspects of competence: *'the advantage of the details in the cards are that they point to aspects that would probably not have been covered otherwise'*. The programme helped in structuring training, for example, gaining access to experience in the intensive care unit. Also the programme fostered teaching and learning through the studying for the tests and through the feedback on performance during the observations.

Table 4. Summary of responses to open-ended questions about positive and negative aspects of an in training assessment programme for first year trainees in anaesthesiology from fifteen trainees. An x indicates number of times a statement was found in the commentaries.

Number of statements	
	Positive statements
16	The programme in general
xxxx	Programme has high quality, well structured and easy to use.
xxxxxx	Programme makes goals and objectives very clear.
xxxxxx	Objectives are reasonable and relevant. Focus on quality of practice. Good focus on 'soft competencies' also.
12	Positive effect on training and teaching
xxxxx	Good instrument for structuring training.
xxxxxxx	Safeguard for training (ex. intensive care) and for supervision.
20	Positive effect on learning
xxxxxxxxx	Ensures effective and efficient learning of both practice and theory.
xxxxxx	Clinical tests motivate study of theory underlying practice. Good coupling of theory and practice.
xxxxx	Written assignments good for reflection on practice. Good opportunity for studying literature in-depth.
7	Visibility and documentation of competence
xxx	Supervisors get to observe you and hear what you can.
xxxx	Nice to document competence and progress. Documentation good for guidance and appraisal.
	Negative statements
4	Tests in practice
xxxx	Supervisors should be better prepared for the programme and the individual tests. Tests are time-consuming if supervisors are not well prepared.
11	Time and planning
xxxxxx	Conflict with service-demands. Needs better planning.
xxxxx	Written assignments time-consuming. Four assignments too much.
6	Negative effect of structure
xxx	Risk of focus on the tests. Might narrow teaching and learning in other aspects.
xx	Too much structure might take away trainee initiative. Written assignments could be a hindrance to engage in research activities due to time-constraints.
x	Difficult to adapt to structure, control and assessment when you are used to be independent.

Finally documentation of competence and monitoring progress were listed as positive aspects. Negative aspects included mainly problems of time-constraints, and supervisors who were ill prepared for the assessment procedures. One trainee stated that *'when the supervisor is not familiar with the content of the competence-card and the theory coupled to the individual tests the assessment procedure takes too long and obstructs efficient service.'* Only one trainee commented that it could be *'difficult to adapt to structure, control and assessment when you are used to being independent'*.

Discussion

The response rate to the survey on the implementation of the programme was high, 25 (96%) of 26 departments, and the results showed that the in-training assessment programme for first year residents in anaesthesiology was widely implemented across the country. Departments with trainees who had completed the entire first year training had implemented most of the training programme and the vast majority of trainees who had only had part of their training were enrolled in the assessment programme. Unfortunately, the trainees' response rate about the evaluation of the programme was only 56% (15/27). However, the responders had a very consistent message about how the programme fostered structuring of the training and had a positive effect on learning. The trainees' ratings of the individual elements of the programme are in accordance with the data from the departments showing that three elements were less popular than others, namely the recording of experience, the survey on communication skills, and the learning portfolio.

The problems with implementing the recording of experience could be due to technical problems. But another problem is that formative evaluation tends to be difficult to implement if there is no statement of quality measured against defined criteria.[16-18] The Cusum-scoring where a quality standard was included tended to be rated higher than the recording of experience. However, this was not statistically significant and the opinion on the Cusum-scoring showed large variances. It has been shown that although formative student logs do provide insight into the nature of students' learning experiences, they do not have the effect of engaging students, teachers or programme directors in structuring training.[19] Logbooks might not be regularly analysed or attended to by trainees and supervisors and are not perceived as a valuable contribution to assessment of competence.[20]

The survey on communication skills was not considered worth the effort without further explanations and this aspect will have to be studied further. The less optimal implementation of the learning portfolio could

be explained by lack of appreciation of the concept. Although the learning portfolios apply to theories of self-directed reflective learning, trainees may find them lacking relevance and not worth the effort.[14, 15] Our highly structured assessment programme, including documentation of procedural performance as well as reflective written assignments, could have contributed to not finding the individual learning portfolios worthwhile.

One concern raised about our programme was that of ill prepared supervisors. The problem of clinicians' preparedness for the assessment protocols and instruments used in in-training assessment has been demonstrated in other programmes.[16, 21] Our results point to the need for better introduction of the supervisors and formal training in assessment protocols; in that respect from the current study, our information meetings and the handbook on assessment appear to have been insufficient. The implementation of the written assignments in general and the trainees' ratings of these were surprisingly high. In a previously conducted nation-wide survey among consultants the test 'evidence-based medicine' was rated significantly lower regarding importance by consultants from departments with first-year trainees compared to other consultants.[14] Due to the awareness of this, special workshops had been run for preparing the supervisors and the trainees and this probably contributed to the better than anticipated implementation.

The summative aspects of the major parts of our programme may well account for the implementation as well as the positive effects on training and learning. Several studies have shown that assessment is a strong incentive to both learners and the clinical departments to engage in learning activities and to structure the training.[22, 23] Time constraints as a barrier to implementation was stated by 5 of the 26 departments and this was less than expected. The programme's effect on structuring training might have been perceived as an advantage that counterbalances the investment of time. A study by Long[24] demonstrated in a neurosurgery residency programme that structuring training through in-training assessment could considerably reduce the time required by trainees to acquire competence. Similar effects of our programme might have played a role in the results of our study. However, this was only vaguely stated in the commentaries and further studies are needed on how the programme works in practice.

Our in-training assessment programme focuses on first-year trainees and on assessing trainees' performance in real life setting. However, other strategies might be necessary for higher levels o f experience and management of more complex situations. Advanced simulators are now available for training and assessment of complex skills and situations

rarely occurring[25, 26] and might be a valuable supplement to in-training strategies.

Conclusion

This study shows that the in-training-assessment programme for first year residents in anaesthesiology has been widely implemented across the country. Preliminary data on the evaluations from trainees indicate that the majority of the programme is highly acceptable and has a positive effect on structuring the training and on promoting learning.

Acknowledgements: This study was supported in part by grants from the Ministry of Health in Denmark. The authors wish to thank all the anaesthetists who took time to fill in the forms and for their willingness to share information about their experiences with the implementation process and for their constructive criticism and suggestions for improvement of the assessment programme. We wish to thank the working group behind the assessment programme for their enthusiasm and efforts in fostering the implementation of the programme. The working group that compiled the curriculum and the assessment programme for first year residency in anaesthesiology included: Andersen N, Berlac PA, Bested K, Callesen T, Christensen P, Jensen E, Jensen JW, Lemholt K, Lund J, Malling B, Mandøe H, Nørregaard O, Pedersen BD, Petersen JA, Ravlo O, Ravn L, Ringsted C, Skjelsager K, Sprehn M, Østergaard D. Finally we wish to thank Deborah Davis for her editorial assistance.

References

1. The Ministry of Health. The future specialist. Report from the specialist commission. Report no 1385. Copenhagen: Statens Information, 2000.
2. The National Board of Health. Guidelines for curriculum of specialist education. The National Board of Health, Copenhagen, 2001.
3. CanMEDS 2000: Extract from the CanMEDS 2000 Project Societal Needs Working Group Report. Med Teach 2000; 22: 549-54.
4. Accreditation Council for Graduate Medical Education. Outcome project, 2000 www.ACGME.org.
5. Friedman Ben-David M. The role of assessment in expanding professional horizons. Med Teach 2000; 22: 472-7.
6. Shepard LA. The role of assessment in a learning culture. Educ Res 2000;29:4-14.
7. Iirola T, Lund VE, Katila AJ, Mattila-Nuori A, Palve H. Teaching hospital physicians' skills and knowledge of resuscitation algorithms are deficient. Acta Anaesthesiol Scand 2002; 46: 1150-4.

8. Skrifvars MB, Castren M, Kuroloa J, Rosenberg PH. In-hospital cardiopulmonary resuscitation. Organization, mangement and training in hospitals of different levels of care. Acta Anaesthesiol Scand 2002; 46: 458-63.
9. Schmidt A, Akeson J. Practice and knowledge of cricoid pressure in southern Sweden. Acta Anaesthesiol Scand 2001; 45: 1210-4.
10. Rosenstock C, Møller J, Hauberg A. Complaints related to respiratory events in anaesthesioa and intensive care medicine from 1994 to 1998 in Denmark. Acta Anaesthesiol Scand 2001; 45: 53-8.
11. Karle H, Nystrup J. Comprehensive evaluation of specialist training: an alternative to Board examinations in Europe? Med Educ 1995; 29: 308-16.
12. Zorab JSM. The educational activities of the European Academy of Anaesthesiology. Minerva Anesthesiologica 1998; 64: 477-82.
13. Ringsted C, Østergaard D, Scherpbier A. Embracing the new paradigm of assessment in residency training. An assessment programme for first year residency training in anaesthesiology. Med Teach 2003; 25: 54-62.
14. Ringsted C, Østergaard D, Scherpbier A. Consultants' opinion on a new practice based assessment programme for first year trainees in anaesthesiology. Acta Anaesthesiol Scand 2002; 46: 1119-23.
15. Kestin IG. A statistical approach to measuring the competence of anaesthetic trainees at practical procedures. Br J of Anaesth 1995; 75: 805-9.
16. Morris ZS, Bullock AD, Belfield CR, Butterfield S, Frame JW. Assessment in postgraduate dental education: an evaluation of strengths and weaknesses. Med Educ 2001; 35: 537-43.
17. Kelly DR, Murray TS. The development and evaluation of a personal learning log for Senior House Officers. Med Educ 1999; 33: 260-6.
18. Mangione S, Peitman SJ. Revisiting physical diagnosis during the medical residency: It is time for a logbook - and more. Acad Med 1999; 74: 467-9.
19. Dolmans D, Schmidt A, van der Beeck J, Beintema M, Gerver WJ. Does a student log provide a means to better structure clinical education? Med Educ 1999; 33: 89-94.
20. Nixon MC. The anaesthetic logbook - a survey. Anaesthesia 2000; 55: 1076-80.
21. Noel GL, Herbers JE, Caplow MP, Cooper GS, Pangaro JN, Harvey J. How well do internal medicine faculty members evaluate the clinical skills of residents? Ann Int Med 1992; 117: 757-65.
22. Petrusa ER. Clinical performance assessments. In: Norman GR, Van der Vleuten CPM, Newble DI, eds. International Handbook of Research in Medical Education. Kluwer Academic Publishers, 2002: 673-709.
23. Van der Vleuten CPM. The assessment of professional competence: developments, research and practical implications. Adv Health Sci Educ, 1996: 1: 41-67.
24. Long DM. Competency-based residency training: the next advance in graduate medical education. Acad Med 2000; 75: 1178-83.
25. Jacobsen J, Lindekaer AL, Ostergaard HT, Nielsen K, Ostergaard D, Laub M, Jensen PF, Johannesen N. Management of anaphylactic shock evaluated using a full-scale anaesthesia simulator. Acta Anaesthesiol Scand 2001; 45: 315-9.
26. i Gardi T, Christensen UC, Jacobsen J, Jensen PF, Ording H. How do anaesthesiologists treat malignant hyperthermia in a full-scale anaesthesia simulator? Acta Anaesthesiol Scand 2001; 45: 1032-5.

CHAPTER 5

Educational impact of ITA

Published in Medical Education (in press)

Educational impact of in-training assessment (ITA) in postgraduate education. A qualitative study of an ITA programme in actual practice.

C Ringsted, A-H Henriksen, AM Skaarup, CPM van der Vleuten

Abstract

Objectives To investigate experiences and thoughts of programme directors, clinical supervisors, and trainees about an in-training assessment (ITA) programme on a broad spectrum of competence for first year training in anaesthesiology: how does the programme work in practice and what are the benefits and barriers; what are the users' experiences and thoughts about the effect on training, teaching and learning; what are their attitudes towards this concept of assessment.
Methods Semi-structured interviews were conducted with programme directors, supervisors, and trainees from three departments. Interviews were audio taped and transcribed. Content of the interviews was analysed in a consensus process among the authors.
Results The programme was of benefit in making goals and objectives clear, in structuring training, teaching and learning, and in monitoring progress and managing problem trainees. There was a general positive attitude towards assessment. Trainees especially appreciated the coupling of theory to practice and in general the programme inspired an academic dialogue. Issues of uncertainties regarding standards of performance and conflict with service declined over time and experience with the programme, and departments tended to solve practical problems through structured planning.
Discussion Three inter-related factors appeared to influence the perceived value of assessment in postgraduate education. 1) The link to patient safety and individual practice using assessment as a licence to practice un-supervised rather than end of training examination, 2) the benefits to educators and learners in the educational process rather than mere documentation of competence, and 3) the attitude and rigour of assessment practice.

Key learning points

- ITA makes goals and objectives clear and can serve as a safeguard for training, teaching and learning.
- Task specific checklists including theoretical aspects are highly appreciated in a postgraduate context.
- ITA should be linked to practice and patient safety with progressive timely assessment instead of end-of-training assessment.
- ITA is of help in identification, diagnosing, and remediation of problem trainees at an early stage.
- ITA instruments are well received by the users as far as they are educational beneficial rather then mere documentation of competence.

Introduction

The current literature on in-training assessment (ITA) indicates the potential benefits to the learner, the teacher, and the programme.[1-5] By emphasizing broader aspects of the future roles of doctors such as communicator, health advocate, collaborator, manager, scholar, and issues of professionalism the development of appropriate ITA models becomes increasingly important.[6] Postgraduate education consists mainly of progression through workplace experience. In developing ITA instruments, the emphasis should be on residents' progression towards defined standards of performance, assessment of professional knowledge as well as assessment of core knowledge, assessment of on-the-job-learning and learning through assessment rather than assessment of learning by the end of training.[3, 7-9]

Medical curricula embracing broader aspects of competence have been described together with the principle of using ITA as an on-going process rather than an end of training activity.[10-15] In order for any assessment programme to be accepted by the users the majority must feel that the process is fair and educationally beneficial.[13, 16] Several studies have dealt with the psychometric properties of ITA and demonstrated that sampling across several observations and using several assessors can ameliorate problems of reliability.[6, 9, 14, 17-19] However, there is currently little research on how ITA programmes in postgraduate education work in practice and what impact they have on the educational process.[7, 20]

The aim of this study was to investigate the experiences and thoughts of the programme directors, the assessors, and the trainees about a recently introduced ITA programme. Firstly we wanted to explore how the programme works in actual practice: how the administration of the ITA programme was organised, how the individual assessments were administered, and what benefits or barriers the users had experienced. Secondly, we wanted to explore what impact the introduction of ITA had on the educational process: what was the effect on training, teaching and learning, and what were the users' experiences and thoughts about the benefits and drawbacks. Finally we wanted to investigate the users' attitudes towards the concept of assessment.

The ITA programme

A working group under the Danish Society of Anaesthesiology and Intensive Care Medicine developed a new nationwide in-training-assessment programme for first year residency of anaesthesiology.[10, 21] The programme included twenty-one individual elements that were sequentially dispersed over the entire first year training (see Table 1).[22] The programme was tailored to the professional development of the trainees

over the year with a focus on clinical skills in the beginning of the programme and on reflection on practice at a later stage of the programme.[23] In the ITA programme standards of performance were elaborated in task specific competence cards that included a checklist and some theoretical questions applying to the procedure (see Table 2). Each card also included a global pass/fail decision on the specific task. The trainee had to pass all elements of the programme in order to have the entire 1st year training approved.

The ITA programme was developed through an internal rational validation process through literature studies, expert group discussions and piloting of parts of the programme.[10] Content validity had been assured by a survey among consultants.[21] Another nation-wide survey revealed that the programme was widely implemented.[24]

Method

This study was performed 1½ year after the introduction of the formal in-training assessment programme. Three departments of anaesthesiology at different hospitals were chosen to participate in the study of which we knew they had implemented the programme. Individual interviews were made with the three programme directors (PD), nine supervisors, and fourteen trainees. The interviews were semi-structured and included questions about the organisation of the ITA programme and how it works in practice, and about positive as well as negative experiences and thoughts of the programme.

Two of the authors (AH and AS) who were not doctors and who had not taken part in the design, development or implementation of the ITA programme performed the interviews. All interviews were audio taped and transcribed. Content was coded and analysed according to the questions and organised into a framework of categories. The general categories include issues of the ITA programme in practice, the educational impact, and the attitude towards assessment. Results are presented according to these general categories with specification of sub-categories. Emergent themes were identified in a consensus process among the first three authors (CR, AH, AS). The team then discussed the summaries of the data and consistency in interpretation.

Table 1 Specification of 21 elements included in the assessment programme for first year trainees in anaesthesiology.

Clinical skills 1. Managing airways 2. Testing anaesthesia machine 3. General anaesthesia 4. Emergency induction of anaesthesia 5. Preoperative consultation 6. Advanced resuscitation 7. Plan for fluid and nutrition 8. Function of standard respirator 9. Managing intensive care patient 10. Spinal anaesthesia 11. Epidural anaesthesia 12. Central venous catheter	These assessments are performed by having a supervisor observe the trainee in practice and score the performance by use of a competency-card with task specific items regarding performance and some questions about theory. Each assessment includes an overall decision regarding pass/fail. The first six elements are supposed to be passed within the first three months of training and before the trainee is assigned to on-call duty.
Experience 13. Recording of experience 14. Cusum-scoring	Recording of experience includes various kinds of patients and procedures. An electronic database was available on the Internet. Cusum-scoring is registration of success or failure rate each time a procedure is performed. Four procedures were planned for Cusum-scoring, spinal anaesthesia, epidural anaesthesia, arterial catheter, and central venous catheter. Success and failure rate is specified taking into account the 'normal' failure rate for expert anaesthetists (see ref. no 22). Electronic database available.
Communication 15. Survey on patient communication	The trainee hands out a questionnaire to 25 patients regarding the preoperative consultation and elaborates a critical report on the results. The supervisor assesses the report and makes a pass/fail decision.
Managerial and collaborative skills 16. Organisation, collaboration, professionalism on call 17. Organisation, collaboration, professionalism in OR	A senior colleague that have worked closely with the trainee rate the trainee on a structured form, one senior colleague rates for assessment no 16, and two for assessment no 17. The trainees' clinical supervisor makes an overall decision regarding pass/fail on behalf of the ratings on each of two assessments.
Written assignments 18. Choice of anaesthetics 19. Reflection on case 20. Evidence based medicine	Written assignments are scored according to specified criteria and an overall pass/fail decision is made for each of them. These assignments are supposed to be placed in the second half of the training period.
Self-directed learning 21. Learning portfolio	The trainee identifies individual learning objectives, means to reach the goals and reports on the learning. The supervisor monitors the progress of the trainee's learning and gives an overall mark on the professional development according to specified criteria on a checklist.

Table 2 Example of checklist used for assessment

Conduct of a spinal anaesthesia
The trainee indicates when he/she is ready for assessment and finds a senior colleague that can assess him/her according to this checklist. All items on the checklist must be correct in order to approve the competence.

	Yes	No
• Outline plan for anaesthesia including description of indication of proper effect, acceptable blood pressure values, plan for managing side-effects or unwanted effects.	☐	☐
• Make proper preparations in the anaesthetic and operating rooms	☐	☐
• Can explain choice of anaesthetic and dose related to the planned surgery	☐	☐
• Inform and instruct patient properly regarding the procedure	☐	☐
• Perform proper sterile preparation; check medicine and instruments	☐	☐
• Make correct positioning of patient in collaboration with assistant	☐	☐
• Indicate correct site of needle prick	☐	☐
• Use proper sterile technique	☐	☐
• Use correct needle technique and assures correct needle position before injecting anaesthetic	☐	☐
• Perform proper observation and monitoring of patient after injection of local anaesthetic	☐	☐
• Initiate proper measures to support cardiovascular function	☐	☐
• Make proper test of the analgesia and indicate segmental level correctly	☐	☐
• Can name indications and contra-indication to spinal anaesthesia	☐	☐
• Can name at least three complications and risks of spinal anaesthesia; can describe precautions regarding prevention and describe proper management of these complications.	☐	☐
This competence is acceptable	☐	☐

Name of supervisor: Name of trainee: Date:

Results

The ITA programme in practice

Key persons and responsibilities

The PDs had the overall responsibility of the programme. The responsibilities of the personal supervisors included introduction of the trainees to the department and the training, guidance and follow-up on the trainees' progress, and examining trainees, at their request, on any of the individual assessments in the programme. Any other senior colleague in the department was also expected to examine trainees at their request. The PDs had introduced and discussed the ITA programme and the assessment protocol with the staff at several meetings within the departments emphasizing the responsibilities that each individual had towards making this programme work. Assessment protocol and practical problems were addressed regularly at staff-meetings. PDs had briefly introduced the personal supervisors to their assignments, but apart from the staff-meetings, they had not discussed the assessment protocols in detail with the supervisors. Most supervisors reported that they had prepared for the assessments and studied the basic textbook of anaesthesiology in order to capture the level of requirements for first year trainees and review topics that were not within their main area of academic interest.

In principle the trainees had the responsibility to indicate when they wanted to be assessed and to find a senior colleague who was willing to do so. Often that was most feasible when on call. The written assignments were prepared by the trainees at home and assessed by the supervisor and the PD together. One department requested oral presentation and discussion rather than writing.

Problems and barriers

A few commented that the ITA programme was extensive, but at the same time that all elements were very relevant. Issues of time and conflict with service demands were addressed by the interviewees to a minor extent. A few supervisors indicated that they found it difficult in the beginning and that more instruction in the assessment protocol would have been appropriate.

[S6] "Time is a problem especially at the start. It's better now that we're more used to it, and we know the standards and so on. And somehow we seem to make it work."

Both supervisors and trainees appreciated that trainees had the initiative to undergo assessment. Some trainees found it rather stressful to find a willing assessor and tended to be reluctant to ask for assessment, which resulted in lagging behind the programme's schedule.

[T7]: "The benefit is that you're forced to be active and be aware of your own progress. The drawback is the barrier to approach a supervisor. Perhaps some more initiative from the supervisors would be good, but it is difficult to strike the balance between taking way our responsibility and on the other hand not feel too pushy."
[T14]: "You're reluctant because you are asking another person for some of his time – and that's easy if it's something you cannot do, but it's difficult to ask for supervision on something that you can do."

Administration of the individual assessments
Some supervisors adhered strictly to the protocol, i.e. observed the trainee, asked the questions, and checked on the card. Others split the individual assessments in a practical and a theoretical part. A few took it rather leniently and checked the card on behalf of a general impression of the trainee. The trainees pointed to the importance that supervisors take it seriously and trainees tended to seek supervisors that were serious about the assessments, showed interest and were knowledgeable.
[T12] "You look for supervisors that you know are knowledgeable and where you get something out of it, a good dialogue or really learn something. Not just marks on a sheet of paper."

Several trainees and supervisors pointed to the importance of linking assessment to patient safety and using the assessments as a licence to engage in service and become more independent.
[T11]:"I didn't get the epidural right and I didn't pass. But it's ridiculous that I did 3 epidurals alone on the same call.... you should not be allowed to do anything on your own before you have passed."

Change in practice over time
Over time and experience with the programme the recognition of some of these practical problems caused departments to schedule several of the first assessments during the day time within the first one or two months of the trainee's posts with specific appointment of assessors. This structured organisation of the programme was highly appreciated by the supervisors and trainees, who had experience with it.
[T3]: It's nice that they have approved that you can do such and such. You get confidence – I know I have been able to do this correctly."

Educational impact
The interviewees found that the ITA programme was good in making goals and objectives very clear and in drawing attention to other aspects of competence than medical expertise.
[PD2]: "The advantage is that suddenly they can see the objectives. There are things that have been verbalised, which previously were tacit."

The effect on training, teaching and learning

Some of the supervisors used the programme in structuring training activities and several of the trainees used the programme for requesting training experience within specific categories of patients or procedures, for example getting allocated to service in intensive care. The programme was of help in planning and focusing teaching and learning and interviewees repeatedly indicated that the programme was a safeguard for training, teaching and learning.

[S3]: It gives you structure and is a useful tool in many aspects. You spend inordinately many resources when things are not structured."

The trainees indicated that the programme made learning more effective and efficient in assuring the breadth in the learning and in reaching the goals in a shorter time.

[T3]: "I would probably have studied the same, but it would have taken more time. It's very easy to see what you shall study next."

[T10]: "I had to pass certain things before I could move on. So I hurried to finish the various competencies and then asked for more and more."

The assessment scenarios fostered clinical teaching and discussion of clinical matters between supervisor and trainee also on issues not included in the checklists. Some supervisors returned later and gave further perspectives on the matter. Several interviewees expressed that the programme inspired an academic dialogue in the department in general and indicated that all learn more.

[PD3]: "It has made life easier – questioning has been legalised. You get closer and you're forced to supervise more. They get the breadth in cognitive knowledge and the assignments require that you dig deeper."

[S8]: "I find it beneficial to all of us. It is a little extra work, but it also gives some feedback and a dialogue in the daily work. So it's very positive for the department in general."

PDs and supervisors reported that the trainees studied more than previous trainees, and the trainees reported that the assessment programme made them study. Some trainees reported that they would study anyhow, but that the programme guided their studies and pointed to some aspects that they would probably not have focused on otherwise. The interviewees appreciated the programme's coupling of theory to practice and the details in the competence cards.

[PD2] "The ITA programme has made the trainees more active, they reflect more and transfer the concepts to other things. They critically appraise various things at our meetings."

[T6]: "Some of the cards make me study things that I would not have otherwise – you get around the stuff very well. It forces me to study and reflect on things at an early stage in my training year."

[T11]: "The best part is not so much the practical part, but the theory in the cards, because – if you do a spinal it's more important that you know how to manage complications. It's important that you study the theory when you are supposed to use it."
Those trainees who had performed the written assignments in the end of the programme found them an enjoyable and valuable learning experience. They appreciated the way they made them go over some theory again and reflect on practice.

Negative aspects of the programme
Some interviewees indicated a potential risk of a highly structured programme in reducing initiatives on both sides and that trainees who had completed the programme could get a sense of emptiness.
[T10]: "When you've finished the whole programme its like the development stops. That's bad."

Both trainees and supervisors emphasised that the assessments should foster learning. For example the communication skills survey was not appreciated as they only got positive reports from the patients and they did not learn from that. Another aspect of this survey and also of the assessment of collaborative skills was the lack of immediate feedback and dialogue with the assessor.
[T11]: "If you just get this information on paper – 'you are good at this or you are not good at this', you can't use it. Perhaps you should make some sessions where you discuss how you act in various situations."

The written assignments were considered time consuming and both supervisors and trainees, who had not yet experienced that part of the programme considered it a big challenge.
[S8]: "I find it very ambitious and I tell the trainees that. The assignments require that you have scholarly competence."

Monitoring progress and managing problem trainees
The programme helped in keeping track of trainees' progress and both trainees and supervisors expressed that the programme had added a meaningful content to the appraisal meetings, which previously or in other programmes tended to be loose talk. The ITA programme made the individual trainees more 'visible' and their individual needs more clear.

The PDs expressed that the programme made it easier to get an overview of the whole group of trainees and their individual progress. All three departments had experienced problem trainees and expressed that the programme was of help in identifying problem-trainees at an earlier stage and in confronting trainees with the problems. Further the assessments were used in diagnosing the problems and in structuring remediation plans.

[PD1]:"Problem trainees become much more visible. We get attention to the problem at an earlier stage, discuss the problem and take steps to remediation."
[PD2] "We had our doubts about one trainee - didn't know whether he had academic deficiencies or an attitude problem. We asked him to take the assessments. The effect was that he became more explicit in his professional communication with clear formulation of plans for anaesthesia or expressing clearly when he was in doubt, so the nurses felt more safe about him."

Attitude towards assessment

General attitude
It was the PDs' impression that there was a general positive attitude towards the assessment programme although one PD had difficulties involving the most senior colleagues in assessment activities.
[PD3] "Education is perceived as a private matter in our country, we are not used to examining doctors. Especially the seniors have some resistance. Of course it's also a test for them."

None of the supervisors expressed negative attitudes towards being assessors and the majority of trainees expressed a positive attitude towards assessment in general.
[T12]: "I don't think you should be afraid of making it an examination situation, and doing assessments progressively is good - then you know where you are getting at."
A couple of trainees and a single supervisor found that assessment should be placed in the end of the training year. A single trainee expressed a general negative attitude towards the concept of assessment.
[T9] "I find that these assessments are offending to my perception of myself as a professional doctor. I have made up my mind that I have studied enough in my life. I find that the assessments on my organisational and collaborative skills are too personal. I don't like that."

Documentation of competence
In some aspects the documentation of competence was seen as an advantage especially by the PDs.
[PD1] "Assessment is an assurance that the trainees who have completed the programme and leave our department are good. We are several doctors that have assessed them and that is nice. It gives me a clear conscience."
[PD3]: "We have better documentation of our decision. It's more professional in this way. It's a national standard, not just my personal opinion."

Only a few trainees indicated an advantage of documentation on their self-confidence. But in other aspects the mere documentation was apparently not an important issue to several of the interviewees.
[T8] "It seems a bit ridiculous when you work together with a colleague who knows what you can and then these cards must be filled in."
[S7]: "I don't find that documentation means anything except that filling in these cards become a goal in itself, not a mean to reach the goals."
Several of the interviewees stressed the importance of making assessment mandatory and making that clear to all parties. That was not always clear to the trainees in the beginning.

Discussion

This study reveals that implementation of an in-training assessment programme is beneficial to the learners, the teachers, and the programme directors. The unanimous expression by all three parties of interviewees was that the ITA programme was of benefit in making goals and objectives clear and in structuring training and learning. Also not surprisingly this study demonstrates that assessment fosters teaching and learning.[9, 25] The data suggest three inter-related factors that appeared to influence the perceived value of assessment:

- Link to practice
- Educational benefits to the users
- Attitude and rigour of assessment

Link to practice
The results indicate that assessment seems more meaningful if linked to practice and applied progressively as a licence to work more independently rather than being placed as an end of training examination. Included in this aspect was the issue of patient safety. Another aspect of linkage to practice was the theoretical questions in the checklist related to the practical procedures and the reflection on practice in the written assignments. This coupling of theory to practice was highly appreciated and motivated both trainees and supervisors to study and inspired an academic dialogue in general. Although there are more reliable methods for assessing cognitive knowledge such as multiple-choice tests, these may not have the same educational impact and acceptance.[9, 24]

Educational benefit
In order for the users to perceive a value of the ITA the relevance to the individual user is of significance[16]. In various ways both trainees and supervisors expressed that the assessment should include a challenge

and have an effect on learning. For instance trainees valued assessments that they learned from even when they did not pass at the first attempt, whereas both trainees and supervisors found the patient communication survey meaningless due to only positive ratings by the patients and hence no effect on learning. Interestingly the mere documentation of competence was not an important issue to either trainees or supervisors and both parties expressed a rather negative point of view on this as 'just marks on a sheet'. Only a few trainees expressed that documentation of competence gave them confidence. Perhaps the issue of confidence would have been more pronounced if assessments were coupled to independent practice. Documentation of competence was not a major issue to the supervisors. Due to working schedules clinical teaching is a shared responsibility among all senior clinicians and the progress of the individual trainee is not attributed to the individual supervisor. Thus supervisors might not profit from the mere documentation, but rather see that as an extra burden. The supervisors though profited from the programme in structuring training and teaching activities. The issue of documentation was more important to the PDs. They expressed a sense of security in having evidence and that documentation was of help in managing problem trainees, which indicate a relevance to the PDs' practice and overall responsibilities.

Attitude and rigor of assessment

During introduction of ITA, resistance to the idea of assessment can be expected because of negative associations with examinations.[26] In our study only a single interviewee expressed a negative attitude to assessment. The major concern about our programme in practice was not whether it was fair but rather the trainees requested more rigorous assessment. Assessor leniency and scale shrinking have been repeatedly demonstrated when using clinicians as assessors.[1, 14, 17, 18] Although assessment instruments using general categories of competence have been shown to have sufficient reliability, they may not be able to demonstrate increasing levels of competence over the first year of training.[14] One approach to that problem is to have the clinician assessors rate the trainees on narrative developmentally-oriented rating scales without indication of trainee's sufficiency, but leaving judgement and pass/fail decisions to an educational committee.[15] On the other hand instruments that do not have summative consequences may have minor or no effect on structuring training and learning[27], whereas an ITA instrument emphasizing sequential mastery of procedures can accelerate the learning process dramatically.[7] In our study there were several indicators of how the mandatory issues of the ITA were a driving force in the effectiveness and efficiency of the learning.

The written assignments were seen as a huge challenge especially to those who had not experienced this part of the programme yet. A negative attitude towards the written assignments could be expected in a speciality where procedures are predominant.[21, 23, 28] However, supervisors have a special role in making explicit the value of engaging in scholarly activities and of motivating trainees in areas in which they do not readily engage.[23, 28] This study emphasise the importance of preparing supervisors for that role.

Limitations to the study

The interviewers (AH, AS) were not familiar with the ITA programme and protocol and data were not processed or discussed along the study. The weakness of this was that emergent themes or misunderstandings that could have come up during the interviews were not probed deeper in subsequent interviews. We chose this disadvantage over the possible bias a content expert (CR) would have imposed on the interviews.

This study was performed 1½ year after the introduction of formal in-training assessment, so the data do not represent a steady state context. However, the data give a first impression of how the organisation learns through the experience of implementing ITA. Issues of conflict with time and service demands decline over time and experience with the programme. The study included three departments that were known to have implemented the ITA programme. Thus the results, especially the rather positive attitude towards assessment might have been biased and cannot be generalized as to the feasibility of implementing the ITA in general. On the other hand this study demonstrates what educational benefits ITA can induce even in departments where education already has a high priority.

Acknowledgements: We wish to thank the interviewees for their time. We are grateful to Deborah Davis for her editorial assistance.

References

1. Spike N, Alexander H, Elliott S, Hazlett C, Kilminster S, Prideaux D, Roberts T. In-training assessment - its potential in enhancing clinical teaching. Med Educ 2000; 34: 858-61.
2. Short JP. The importance of strong evaluation standards and procedures in training residents. Acad Med 1993; 68: 522-5.
3. Friedman Ben-David M. The role of assessment in expanding professional horizons. Med Teach 2000: 22: 472-7.
4. Turnbull J, Gray J, MacFadyen J. Improving in-training evaluation programs. J Gen Int Med 1998; 13: 317-23.

5. Feletti G, Cameron D, Dawson-Saunders B, Des Groseilliers J, Dooley B, Farmer E, McAvoy P. In-training assessment. In Newble, D, Jolly, B, Wakeford, R, eds. The certification and recertification of doctors: Issues in the assessment of clinical competence. Cambridge University Press. 1994: 151-66.
6. Turnbull J, Van Barneveld C. Assessment of clinical performance: In-training evaluation. In: Norman, GR, Van der Vleuten, CPM, Newble, DI, eds. International Handbook of Research in Medical Education, Dordrecht: Kluwer Academic Publishers. Printed in Great Britain. 2002 :793-810.
7. Long DM. Competency-based residency training: the next advance in graduate medical education. Acad Med 2000; 75: 1178-83.
8. Miller GE. The assessment of clinical skills/competence/performance. Acad Med 1990; 65(Suppl): S63-7.
9. Van der Vleuten CPM. The assessment of professional competence: developments, research and practical implications. Adv Health Sci Educ, 1996: 1: 41-67.
10. Ringsted C, Østergaard D, Scherpbier A. Embracing the new paradigm of assessment in residency training. An assessment programme for first year residency training in anaesthesiology. Med Teach 2003; 25: 54-62.
11. Reisdorff EJ, Hayes OW, Carlson DJ, Walker GL. Assessing the new general competencies for resident education: A model from an emergency medicine program. Acad Med 2001; 76: 753-7.
12. Prescott LE, Norcini JJ, McKinlay P, Rennie JS. Facing the challenges of competency-based assessment of postgraduate dental training: Longitudinal evaluation of performance (LEP). Med Educ 2002; 36: 92-7.
13. Hays R, Wellard R. In-training assessment in postgraduate training for general practice. Med Educ 1998; 32: 507-13.
14. Durning SJ, Lannie JC, Markert RJ, Pangaro LN. Assessing the reliability and validity of the mini-clinical evaluation exercise for internal medicine residency training. Acad Med 2002; 77: 900-4.
15. Yukowsky R. Can resident evaluations demonstrate increases in residents' skills over time? Acad Med 1999; 74 (suppl): S108-10.
16. Finucane PM, Barron SR, Davies HA, Hadfield-Jones RS, Kaigas TM. Towards an acceptance of performance assessment. Med Educ 2002; 36: 959-64.
17. Noel GL, Herbers JE, Caplow MP, Cooper GS, Pangaro JN, Harvey J. How well do internal medicine faculty members evaluate the clinical skills of residents? Ann Int Med 1992; 117: 757-65.
18. Kroboth FJ, Hanusa BH, Parker S, Coulehan JL, Kapoor WN, Brown FH, Karpf M, Levey GS. The inter-rater reliability and internal consistency of a clinical evaluation exercise. J Gen Int Med 1992; 7: 174-9.
19. Turnbull F, MacFadyen J, van Barneveld C, Norman G. Clinical work sampling. A new approach to the problem of in-training evaluation. J Gen Intern Med 2000; 15: 556-61.
20. Patel BS, Feerick A. Will competency assessment improve the training and skills of the trainee anaesthetist? Anaesthesia 2002; 57: 710-31.
21. Ringsted C, Østergaard D, Scherpbier A. Consultants' opinion on a new practice based assessment programme for first year trainees in anaesthesiology. Acta Anaesthesiol Scand 2002; 46: 1119-23.
22. Kestin IG. A statistical approach to measuring the competence of anaesthetic trainees at practical procedures. Br J Anaesth 1995; 75: 805-9.
23. Rasmussen J. Skills, rules, knowledge, signals, signs and symbols and other distinctions in human performance models. IEEE Trans Systems, Man and Cybernetics 1983; 123: 257-66.
24. Ringsted C, Østergaard D, Van der Vleuten CPM. Implementation of formal in-training assessment in anaesthesiology and preliminary results of acceptability Acta Anaesthesiol Scand 2003; 47: 1196-1203.

25. Petrusa ER. Clinical performance assessments. In: Norman, GR, Van der Vleuten, CPM, Newble, DI, eds. International Handbook of Research in Medical Education. Kluwer Academic Publishers, 2002: 673-709.
26. McManus IC, Winder BC, Gordon D. UK doctors' attitudes to the General Medical Council's performance procedures, 1997-99. Med Educ 2001; 35(Suppl): 60-9.
27. Dolmans D, Schmidt A, van der Beck J, Beintema M, Gerver WJ. Does a student log provide a means to better structure clinical education? Med Educ 1999; 33: 89-94.
28. Baker JD, Cooke JE, Contoy JM, Bromley HR, Hollon MF, Alpert CC Beyond career choice: the role of learning style analysis in residency training. Med Educ 1988; 22: 527-32.

CHAPTER 6

Effect of ITA on clinical confidence

Published in Medical Education 2004 (in press)

The effect on in-training assessment on clinical confidence in postgraduate education.

C Ringsted, J Pallisgaard, D Østergaard, A Scherpbier

Abstract

Introduction The literature on how in-training assessment (ITA) works in practice and what educational outcomes can actually be achieved is limited. One of the aims of introducing ITA is to increase trainees' clinical confidence based on the assumption that assessment drives learning through its content, format and programming. The aim of this study was to investigate the effect of introducing a structured ITA programme on junior doctors' clinical confidence. The programme was aimed at first year trainees in anaesthesiology.
Methods The study was a nation-wide survey of junior doctors' self-confidence in clinical performance before (year 2001) and two years after (year 2003) the introduction of an ITA programme. Respondents indicated confidence on a 155-item questionnaire related to performance of clinical skills and tasks reflecting broad aspects of competence. Twenty-three of these items related to the ITA programme.
Results Response rate was 377/531 (71%) in year 2001 and 344/521 (66%) in year 2003. There were no statistically significant differences between mean level of confidence before and two years after the introduction of the ITA programme – neither in aspects related to the programme nor in aspects not related to the programme.
Discussion This study demonstrates that the introduction of a structured ITA programme did not have any significant effect on the trainees' mean confidence on a broad range of aspects of clinical competence. The importance of timeliness and rigorousness in the application of ITA is discussed.

Introduction

Many authors have described the potential of in-training assessment (ITA) in postgraduate education and a wide variety of assessment instruments for comprehensive ITA programmes have been suggested.[1-4] Trainees in postgraduate education want ongoing supervision and feedback and on-the-job discussions with consultants.[5] They want algorithms for medical decision-making, regular sign-offs in a competencies logbook, effective and strategic use of annual assessment, and development of more objective instruments for use in psychomotor skills training and assessment.[5] However, the literature on how ITA programmes work in practice and what educational outcomes can actually be achieved is limited.[6]

Overview

What is already known on this subject
Many authors have described the potential benefits of in-training assessment (ITA) in post-graduate education is limited.

What this study adds
This study demonstrates no difference in mean level of clinical confidence in nation-wide survey of trainees before and two year after the introduction of formal ITA. Timeliness of ITA related to trainees' clinical experience might influence the effect of ITA on confidence.

Suggestions for further research
In many specialties application of the principles of sequenced ITA is still in its infancy and the full effect of ITA on the learning process and outcomes remains to be explored in future research.

Assessment drives learning through its content, format and programming.[7] Introducing assessment in programme aspects that have not previously been part of the assessment curriculum might increase both trainees' levels of competence and change their attitude towards the importance of those aspects.[8] Similarly, introducing assessment might change clinical teachers' practice such that they provide increased supervision in those aspects included in formal assessment.[9] One of the aims of introducing ITA is to increase trainees' clinical confidence based on the assumption that knowledge of actual competence gained through assessment increases self-confidence.[10] Timeliness of ITA related to clinical experience could be expected to influence the perceived effect on learning. In a previous study on how a specific ITA programme works in actual practice the users recommended that the individual assessments be applied in a timely fashion and used as a licence to practice more independently rather than be placed at a late stage of the training year.[11] Long demonstrated that sequenced ITA considerably reduced training time needed for mastery of procedures.[12]

The aim of this study was to investigate the effect of introducing a structured ITA programme on junior doctors' clinical confidence. The first research question was: What is the junior doctors' self-confidence in performing a broad range of clinical skills and tasks before and two years after the introduction of a structured ITA programme – both in aspects related to the ITA programme and in aspects not related to the programme? The second research question was: Has the ITA programme been applied on a timely basis according to the assessment protocol? The third research question was: To what extent does ITA contribute to trainees' self-confidence relative to clinical experience?

Context of the study

A working group under the Danish Society of Anaesthesiology developed a new nationwide ITA programme for first year trainees in anaesthesiology.[10] The programme was introduced in the spring of 2001. Previously there had been no formal assessment or examination of trainees in Denmark. The ITA programme was developed through an internal rational validation process using literature studies, expert group discussions and piloting of parts of the programme. The programme included twenty-one individual elements addressing aspects of medical expertise as well as general aspects of competence.

Based on the internal rational validation process an assessment protocol was elaborated which recommended that the programme be sequentially dispersed over the entire first year of training with a focus on clinical skills in the beginning of the training year and other aspects of competence at a later stage of the year (see Table 1).

Standards of performance were elaborated in task specific competence cards that included a checklist and some theoretical questions about the procedure. Each card also included a global pass/fail decision on the specific task. The trainee had to pass all elements of the programme in order to have the entire 1st year training approved. These standards and the assessment protocol were distributed to all clinical departments of anaesthesiology. Content validity and agreement on the assessment protocol had been assured by a nation-wide survey among consultants.[13] In order to prepare the staff for conducting ITA a handbook including instructions to assessors were distributed to all clinicians and a number of seminars were offered across the country by the introduction of the ITA programme. Widespread implementation of the ITA programme was documented through a nation-wide survey 1½ year after the introduction of the programme.[14]

Table 1 Specifications for the 21 elements included in the assessment programme for first year trainees in anaesthesiology.

Clinical skills 1. Managing airways 2. Testing anaesthesia machine 3. General anaesthesia 4. Emergency induction of anaesthesia 5. Preoperative consultation 6. Advanced resuscitation 7. Plan for fluid and nutrition 8. Function of standard respirator 9. Managing intensive care patient 10. Spinal anaesthesia 11. Epidural anaesthesia 12. Central venous catheter	These assessments are performed by having a supervisor observe the trainee in practice and score the performance by use of a competency-card with task specific items regarding performance and some questions about theory. Each assessment includes an overall decision regarding pass/fail. The first six elements are supposed to be passed within the first three months of training and before the trainee is assigned to on-call duty.
Experience 13. Recording of experience 14. Cusum-scoring	Recording of experience includes various kinds of patients and procedures. Cusum-scoring is registration of success or failure rate each time a procedure is performed. Four procedures were planned for Cusum-scoring, spinal anaesthesia, epidural anaesthesia, arterial catheter, and central venous catheter.
Communication 15. Survey on patient communication	The trainee hands out a questionnaire to 25 patients regarding the preoperative consultation and elaborates a critical report on the results. The supervisor assesses the report and makes a pass/fail decision.
Managerial and collaborative skills 16. Organisation, collaboration, professionalism on call 17. Organisation, collaboration, professionalism in OR	A senior colleague rates the trainee on a structured form, one senior colleague rates for assessment no 16, and two for assessment no 17. An overall decision is made regarding pass/fail on behalf of the ratings on each of two assessments.
Academic aspects 18. Choice of anaesthetics 19. Reflection on case 20. Evidence based medicine	Trainees' written reports are scored according to specified criteria and an overall pass/fail decision is made for each of them. These assignments are supposed to be carried out in the second half of the training period.
Self-directed learning 21. Learning portfolio	The trainee identifies individual learning objectives, means to reach the goals and reports on the learning. The supervisor monitors the progress of the trainee's learning and gives an overall mark on the professional development according to specified criteria on a checklist.

Methods

This study was a nation-wide cross-sectional survey of self-confidence in clinical performance over the entire cohort of junior doctors in anaesthesiology before, n=531 (in April, year 2001), and two years after, n=521 (in April, year 2003), the introduction of a nation-wide ITA programme. A cross-sectional design was chosen, as introduction-year training programmes do not have a fixed entry term, but varies considerably across the country.

Subjects
The subjects in the survey included five groups of junior doctors: a group of miscellaneous house staff, three groups of trainees in formal training positions, and a group of specialists in anaesthesiology who had not yet been awarded positions as consultants (hence the classification as 'junior doctor'). The three groups of trainees in formal training related to three types of positions: 1) an introduction year, 2) a two-years residency including rotations pertinent to anaesthesiology in various surgical subspecialties, intensive care medicine, pain clinic, etc., and 3) a 1½ years senior residency. The group of miscellaneous house staff was a rather heterogeneous group including doctors waiting for formal training positions, either the introduction year or the residency of two years, and doctors in vacancies or short-term appointments used as supplement for other specialities. The ITA programme was aimed toward the introduction year trainees. However, in 2003, two years after the introduction of the programme, experience with the programme by doctors among the miscellaneous house staff and some of the residents were expected as well. The most senior groups of junior doctors were included in the survey in order to assure construct validity of the questionnaire.

Instrument
A questionnaire on self-confidence in clinical performance was constructed. The questionnaire included 155 items related to a wide range of clinical skills and other aspects of competence. The list of items was selected by a small group of consultants and reviewed by the working group responsible for the ITA programme for the 1st year training. The list represented the objectives for the entire specialist education as issued in year 2000 by the Danish Society of Anaesthesiology and was supplemented by items related to other aspects of competence, which at the time of the study were not defined in the objectives. The 155 items related to four categories of clinical competence: 1) 55 items on technical procedures, 2) 48 items on anaesthesia related to various types of surgery, 3) 30 items related to management of patient cases such as intensive care or

emergency care, and 4) 26 items on other aspects of competence such as difficult communications, organisation and management of operation lists, academic competence, and professionalism. Twenty-three of the items (15%) related specifically to the ITA-programme.

The respondents were asked to indicate their confidence on each item on a five point scale, 1-5, where 1=not at all confident in performing this skill or task and 5=totally confident. For clinical skills 1=not at all confident was defined as 'I always need help, advice, support or supervision from a more experienced colleague' and 5=totally confident as 'I never need help, advice, support or supervision from a more experienced colleague.[15] For the other aspects of competence 1=not at all confident was defined as 'I find this difficult and I have to consciously work at it', and 5=totally confident as 'This is so inherent in my practice that I do it easily and instinctively'. [15] Also included were questions about type of training position, time served in anaesthesiology, and gender. Finally the questionnaire in 2003 included questions on which of the 21 elements in the ITA programme the respondent had experienced as a trainee.

Data analysis

Distribution of respondents on types of training positions and gender was explored using Chi^2-analysis, comparing the data from year 2001 and year 2003. For each respondent a mean level of confidence was calculated separately for those 23 items related to the ITA programme and for those 132 items, which were not related to the programme. Students t-test was used for comparing differences in confidence level within and between groups. Bonferroni correction was used to adjust probability level according to multiple comparisons and $p<0.002$ was required for statistical significant difference at the 0.05 level.

In order to estimate the timeliness of application of ITA, number of assessments performed by the two youngest groups was calculated in time intervals of three months. According to the assessment protocol the first six assessment should be performed within the first three months and the rest of the 21 assessments should be progressively applied over the rest of the first training year. Confidence level was expected to increase according to increasing level of experience in anaesthesiology. The third research question, the contribution of ITA on confidence, was investigated by doing a stepwise linear regression analysis using self-confidence as a dependent measure and time served in anaesthesiology and number of assessments performed as independent variables.

Results

The response rate was 377/531 (71%) in year 2001 and 344/521 (66%) in year 2003. Three respondents did not indicate position and sixteen indicated having other positions than being specialists, trainees or house staff. These nineteen respondents were omitted from further analysis. About half of the respondents in both years were specialists in anaesthesiology. The distribution of respondents from different training positions was similar in the two years (Chi^2-analysis, p=0.300), see Table 2. The ratio men/women within the groups was not statistically significant different in year 2001 compared to year 2003 (Chi^2-analyses, p>0.100 in all comparisons).

Table 2. Distribution of junior doctors in anaesthesiology who responded to a survey on clinical confidence and experience before (year 2001) and two years after (year 2003) the introduction of a structured in-training assessment programme for 1st year trainees.

	2001	Men	Women	2003	Men	Women
Total	377 (100%)	205	160	344 (100%)	181	152
Specialist	187 (50%)	116	71	164 (48%)	96	68
Senior residents	50 (13%)	22	28	38 (11%)	22	16
Residents	53 (14%)*	28	24	49 (14%)	28	21
1st year trainee	55 (15%)	28	27	47 (14%)	17	30
House staff	21 (6%)	11	10	35 (10%)	18	17

*One resident did not indicate gender.

The mean number of years served in anaesthesiology and the mean level of confidence for the 23 items related to the ITA programme and for the 132 items not related to the programme is shown in Table 3 for each of the five groups of junior doctors. There were no statistically significant differences in comparisons of mean level of confidence between year 2001 and 2003 within any of the groups. In both years there was a statistically significant increase in mean level of confidence related to increasing level of formal training positions, the introduction year, residency, senior residency, and specialists. In the four groups of trainees there were no significant differences in mean confidence between men and women, neither in year 2001 nor in year 2003 (Students t-test, p>0.060 in all com-

parisons). In the group of specialists there was a small but significant difference between men and women in year 2001, mean confidence on items related to ITA 4.8 (0.2) vs. 4.6 (0.3), p=0.002, and items not related to ITA, 4.2 (9.4) vs. 4.0 (0.4), p=0.002. In year 2003 there were no significant differences in confidence levels between men and women.

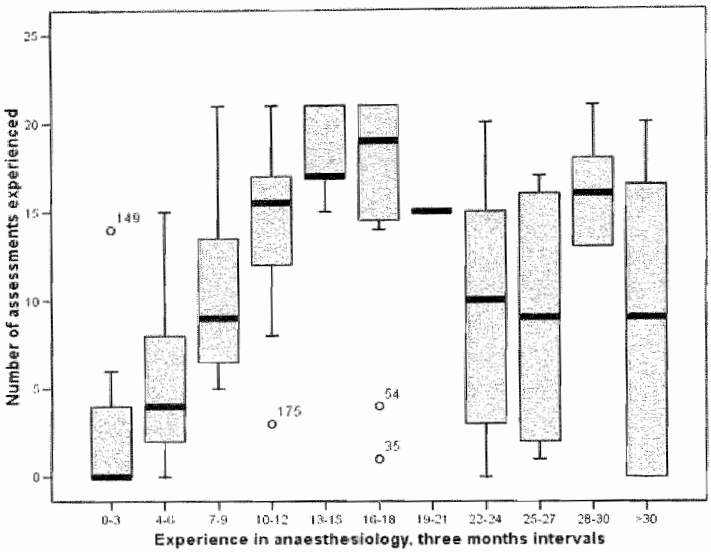
Table 3. Five groups of junior doctors' clinical confidence before (year 2001) and two years after (year 2003) the introduction of a structured ITA-programme for 1st year trainees in anaesthesiology. Confidence was indicated on 155 items related to aspects of clinical competence and indicated on a scale 1-5.

	Mean number of years served in anaesthesiology	Mean level of confidence on items related to ITA (n=23)	Mean level of confidence on items not related to ITA (n=132)
House staff			
Before, N=21	1.6(1.0)	3.6(0.9)	2.4(0.5)
After, N=35	1.2(0.9)	3.3(1.0)	2.4(0.7)
	p=0.267	p=0.207	p=0.613
Introduction year			
Before, N=55	0.8(0.6)	3.4(0.6)	2.2(0.4)
After, N=47	0.8(0.6)	3.6(0.8)	2.4(0.6)
	p=0.768	p=0.283	p=0.188
Residents			
Before, N=53	4.3(2.5)	4.2(0.4)	3.3(0.5)
After, N=48	3.9(1.7)	4.2(0.4)	3.2(0.5)
	p=0.363	p=0.903	p=0.712
Senior residents			
Before, N=50	6.3(1.5)	4.5(0.3)	3.8(0.4)
After, N=38	6.2(2.3)	4.5(0.3)	3.8(0.4)
	p=0.864	p=0.935	p=0.481
Specialists			
Before, N=187	11.9(4.1)	4.7(0.3)	4.2(0.4)
After, N=161	11.8(4.0)	4.7(0.3)	4.1(0.4)
	p=0.763	p=0.108	p=0.427

In year 2003 twenty-seven (77%) of the house staff, forty (85%) of the introduction-year trainees, and 31 (63%) residents reported having experienced one or more of the 21 elements included in the ITA programme. In these three groups of junior doctors the trainees' experience

with the ITA-programme increased according to time served in anaesthesiology over the first 1½ year, Figure 1. However, there were large variances in number of assessments that the trainees had experienced and among trainees having served more than 1½ years the lower range approached zero.

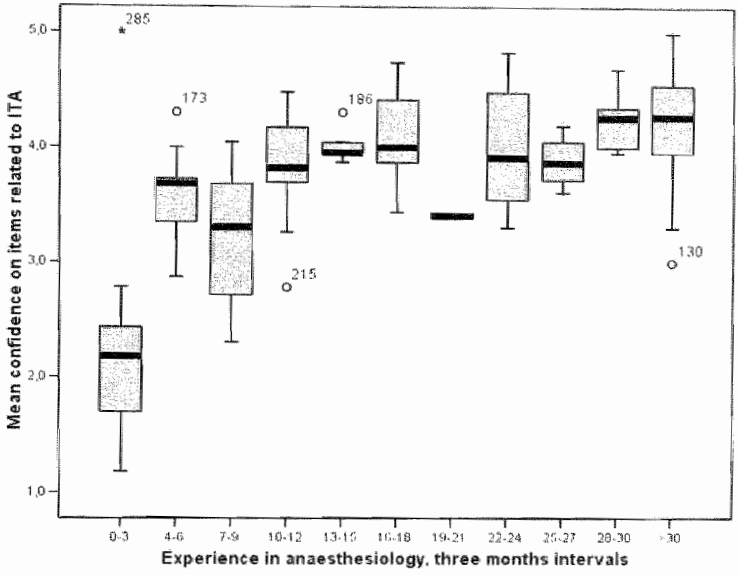
Figure 1. Experience with the 21 elements of the ITA-programme according to time served in anaesthesiology among 131 trainees representing house staff, introduction year trainees, and residents. For each three months interval the box plots demonstrate the median (heavy line) number of the 21 elements applied, the inter-quartile range (the box) and the range of responses (the whiskers).



Mean confidence in items related to ITA increased over the first 1½ year and then seem to level off, Figure 2. For the 67 trainees having served up to 1½ year in anaesthesiology mean confidence in items related to ITA correlated significantly to both time in anaesthesiology, r=0.772, p=0.000, and number of assessments, r=0.671, p=0.000. This correlation was calculated without the outlier, case no 285, see Figure 2. A stepwise linear regression analysis demonstrated a significant contribution of time on mean confidence, R=0.772, R^2=0.595, F=95.654, p=0.000, with a standardized beta coefficient of 0.772, p=0.000. Number of assessments performed did not contribute significantly to the model, R^2change=0.023, Fchange=3.786, p=0.056. Those 63 trainees who had served more than 1½ year in anaesthesiology had experienced median 12 (inter-quartile range 0-17) of the 21 elements of the ITA programme. There was no significant

difference between those who had experienced from zero to three assessments, n=25, and the rest who had experienced from six to 21, n=38, in mean confidence on items related to ITA, 4.1 (SD 0.4) vs. 4.2 (SD 0.5), p=0.422, or items not related to the programme, 3.2 (SD 0.5) vs. 3.2 (SD 0.4), p=0.950.

Figure 2. Mean confidence on items related to ITA according to time served in anaesthesiology among 131 trainees representing house staff, introduction year trainees, and residents. For each three months interval the box plots demonstrate the median (heavy line) of mean confidence levels, the inter-quartile range (the box) and the range of responses (the whiskers).



Discussion

This study is a survey of junior doctors' confidence in performance of a broad range of clinical skills and tasks related to other aspects of competence before and after the introduction of a structured ITA programme. There was an equally high response rate in 2001 and in 2003 with no difference in distribution of respondents from different training positions, which supports the validity of comparisons between the two years. Construct validity of the 155 items surveyed was supported by the observed increase in mean level of confidence over increasing amounts of experience in anaesthesiology. Implementation of the ITA programme was

confirmed by a rather high percentage of trainees reporting experience with the programme.

In previous studies about this ITA, programme directors, supervisors, and trainees have expressed that the ITA had a positive effect on learning and teaching and motivated all to study more.[11, 14] However, the results of this study demonstrate that the introduction of a structured ITA programme did not have any significant effect on the junior doctors' mean level of confidence, neither in aspects related to the programme nor in other aspects.

There are several possible explanations for this result. Assessment by direct observation according to standards of performance could make the trainees more realistic about their actual competence. That could lead to a relatively lower rating of confidence counterbalancing a positive effect of ITA on learning.[16] It has repeatedly been demonstrated that individuals' self-assessment of competence is severely flawed and trainees might feel confident without being competent and vice versa.[16-19] However, at the group level we saw differences in accordance with level of experience in anaesthesiology.

The users of our ITA programme have especially praised the linking of theory to practice and the effect of making the trainees study more.[11, 14] However, an effect of ITA on learning within cognitive aspects is probably not reflected to the ratings of confidence in clinical practice.[18-19] We defined 'confidence' according to the perceived need for support in performing a task. This concept of confidence is not synonymous with the level of competence representing the individuals' actual knowledge and skills related to the task.[15] Confidence represents an individual's readiness to take on a task including a risk, and that is related to previous experience rather than known levels of competence.[15] Unfortunately at the time of the study we did not have any data on individuals' actual knowledge and skills. Future studies including these aspects will be necessary to estimate the full effect of ITA on trainees' learning.

The assessors' rigour probably affects the contribution of assessment to the trainees' learning and clinical confidence.[11] Leniency of clinician assessors is a major problem in ITA[20-23] and trainees express that they don't learn from assessments that are not rigorous.[11] The results might also reflect the clinical departments' inexperience with formal assessment and non-compliance with the timely sequencing of assessments according to the protocol[11, 14] as we saw large variances in number of the 21 elements applied in the first 1½-year. In a previous study on how the ITA programme works in actual practice the users recommended that the individual assessments be applied in a timely fashion and used as a licence to practice more independently rather than be placed at a late stage of the training year.[11] Timeliness of application of the individual assess-

ments might have influenced the contribution of assessment to trainees' confidence relative to the time served in anaesthesiology. Learners are better at recognising development of new skills, but seem less able to identify improvement in existing skills.[16] Thus assessments applied in the beginning as a licence to practice might contribute more to confidence than assessments applied at a late stage of the trainees' professional development.

In our previous studies a few users raised concerns about a risk of taking away responsibility for teaching and learning in aspects not covered by the programme.[11, 14] Any assessment programme can only assess a small sample of the objectives of the entire curriculum and the downside of introducing any structured assessment might be less emphasis on the other aspects of the entire curriculum. However, this potential risk was not confirmed by this study, as the confidence in areas not related to the programme did not change.

In many specialties application of the principles of ITA is still in its infancy and the full effect on the learning process and outcomes remains to be explored. The context of this study was a country with no assessment tradition in postgraduate education and perhaps this study was performed too early after the introduction of ITA. Cultural changes take a long time and there are reasons to believe that users learn over time and experience with ITA.[11]

Conclusion

This study demonstrates that the introduction of a structured ITA programme did not have any significant effect on the trainees' mean confidence as measured in this study on a broad range of aspects of clinical competence - neither in aspects related to the ITA programme nor in other aspects. Timeliness and rigorousness in the application of ITA probably contribute to the effect on clinical confidence.

Acknowledgements: The authors wish to thank the numerous anaesthesiologists who took time to answer the questionnaires. We are grateful to D Davis for her editorial assistance and to H Panduro, S Findalen, and A Schmidt for their help with managing the surveys and the data.

Funding
This study was funded by CHC Postgraduate Medical Institute and in part by grants from the Ministry of Health and from the Danish Society of Anaesthesiology and Intensive Care Medicine. There was no conflict of interest.

References

1. Epstein RM, Hundert EM. Defining and assessing professional competence. JAMA 2002; 287: 226-35.
2. Murray E, Gruppen L, Catton P, Hays R, Woolliscroft JO. The accountability of clinical education: its definition and assessment. Med Educ 2000; 34: 871-9.
3. Turnbull J, Van Barneveld C. Assessment of clinical performance: In-training evaluation. In: Norman, GR, Van der Vleuten, CPM, Newble, DI, eds. International Handbook of Research in Medical Education, Dordrecht: Kluwer Academic Publishers. Printed in Great Britain. 2002; 793-810.
4. Friedman Ben-David M. The role of assessment in expanding professional horizons. Med Teach 2000; 22: 472-7.
5. Khera N, Stroobant J, Primhak RA, Gupta R, Davies H. Training the ideal hospital doctor: the specialist registrars' perspective. Med Educ 2001;35: 957-66.
6. Spike N, Alexander H, Elliott S, Hazlett C, Kilminster S, Prideaux D, Roberts T. In-training assessment - its potential in enhancing clinical teaching. Med Educ 2000;34:858-61.
7. Van der Vleuten CPM. The assessment of professional competence: developments, research and practical implications. Adv Health Sci Educ 1996; 1: 41-67.
8. Newble DI, Jaeger K. The effect of assessment and examinations on the learning of medical students. Med Educ 1983; 17: 165-71.
9. Stillman PL, Haley H-L, Regan MB, Philbin MM. Positive effects of a clinical performance assessment program. Acad Med 1991; 8: 481-3.
10. Ringsted C, Østergaard D, Scherpbier A. Embracing the new paradigm of assessment in resident education. An assessment programme for first year residency training in anaesthesiology. Med Teach 2003; 25: 54-62.
11. Ringsted C, Henriksen AH, Skaarup AM, Van der Vleuten CPM. Educational impact of in-training assessment (ITA) in postgraduate medical education. A qualitative study of an ITA programme in actual practice. Med Educ (in press).
12. Long DM. Competency-based residency training: the next advance in graduate medical education. Acad Med 2000; 75: 1178-83.
13. Ringsted C, Østergaard D, Scherpbier A. Consultants' opinion on a new practice based assessment programme for fist year residents in anaesthesiology. Acta Anaesthesiol Scan 2002; 46: 1119-23.
14. Ringsted C, Østergaard D, Van der Vleuten CPM. Implementation of a formal in-training assessment programme in anaesthesiology and preliminary results of acceptability. Acta Anaesthesiol Scan 2003; 47: 1196-203.
15. Stewart J, O'Halloran C, Barton JR, Singleton SJ, Harrigan P, Spencer J. Clarifying the concepts of confidence and competence to produce appropriate self-evaluation measurement scales. Med Educ 2000; 34: 903-9.
16. Whittle SR, Murdoch-Eaton DG. Student-selected projects: can they enhance lifelong learning skills? Med Teach 2002; 24: 41-4.
17. Mavis B. Self-efficacy and OSCE performance among second year medical students. Adv Health Sci Educ 2001; 6: 93-102.
18. Arnold L, Willoughby TL, Calkins EV. Self-evaluation in undergraduate medical education: A longitudinal perspective. J Med Educ 1985; 60: 21-8.
19. Kegel-Flom P. Predicting supervisor, peer, and self ratings of intern performance. J Med Educ 1975; 50: 812-5.
20. Ringsted C, Østergaard D, Ravn L, Pedersen JA, Berlac PA, Van der Vleuten CPM. A feasibility study comparing checklists and global rating forms to assess resident performance in clinical skills. Med Teach 2003; 25: 654-8.

21. Noel GL, Herbers JE, Caplow MP, Cooper GS, Pangaro JN, Harvey J. How well do internal medicine faculty members evaluate the clinical skills of residents? Ann Int Med 1992; 117: 757-65.
22. Durning SJ, Cation LJ, Markert RJ, Pangaro LN. Assessing the reliability and validity of the Mini-clinical evaluation exercise for internal medicine residency training. Acad Med 2002; 77: 900-4.
23. Norcini JJ, Blank LL, Arnold GK, Kimball HR. Examiner differences in the Mini-Cex. Adv Health Sci Educ 1997; 2: 27-33.

Conclusion and discussion

This chapter is a discussion and conclusion according to the research questions. The overall research question of this thesis was. "How do theory on education and previous research outcomes inform the design and content of ITA in a work-based postgraduate medical education context and how does that apply to a specific ITA programme in actual practice?"

Six sub-questions were generated from the overall research question and applied to the specific context of 1st year training in anaesthesiology. The first part of this chapter is a review and conclusion of the studies on the specific ITA programme reported in this thesis. The second part is a discussion of some general methodological aspects and practical issues related to the studies. The third part is a reflection on the implications for ITA in a wider context of postgraduate education. Finally, an overall conclusion is drawn and some future research questions are outlined.

Review and conclusion of the studies in this thesis

This section is a summary of the findings according to the first research question and a review of the five subsequent studies as they apply to the theoretical considerations and actual outline of a specific ITA programme aimed for 1st year trainees in anaesthesiology.

Research question 1: What is the internal rational validity of ITA in postgraduate education? How do theory on education and previous research outcomes inform the design and content of a specific ITA programme?
Chapter 1 describes how the theory on education and previous research were used to design a specific ITA programme for 1st year trainees in anaesthesiology. The programme portrayed broad aspects of competence and included 21 elements of which 18 were summative assessments and three served a formative purpose. The resulting design recommended a sequential dispersion of ITA over the training year with emphasis on clinical skills in the beginning of the year and other aspects of competence and reflection on practice at a later stage of the year. Further, the validation process resulted in choosing task specific checklists over global scoring forms addressing general aspects of competence based on the assumption that checklists would provide better accuracy in scoring trainees and that they might serve as a help to the assessors in clarifying expectations of performance. The internal rational validation process contributed to foreseeing problems with implementation of ITA in a work-based busy clinical context and the programme was narrowed down as much as possible. The selected content were elements perceived to be of major importance. Problems of acceptance of the elements of the programme addressing academic competence were expected. Finally it was hypothesised that ITA would be of help to trainers and trainees in structuring the teaching, training, and learning and have a positive impact on the trainees' learning and clinical confidence. The appropriateness of the design and content, the feasibility and educational impact was explored in five subsequent studies.

Research question 2: What is the content validity of ITA addressing broad aspects of competence? Do clinical experts agree on the composite content and the pre-defined standards for making decisions about the trainees?
In order to ensure coverage of the curriculum and agreement on the assessment protocol, consensus on the composite content and the predefined standards for making decisions about the trainees was sought through a nation-wide survey among consultants in anaesthesiology, Chapter 2. Based on responses from 251 (66%) of the entire population of consultants we found that more than 75% strongly agreed on the repre-

sentativeness of the assessments included in the ITA programme and the appropriateness of the composite programme for making pass/fail decisions. Further, 75% of the respondents gave high ratings of the comprehensiveness of the content in each of the summative assessments included in the programme. The ratings on the importance of two assessments were slightly lower compared to the others. One was a written assignment on academic competence and the other was an assessment reflecting organisational skills and professionalism on-call. Problems with acceptance of the written assignment were foreseen in the internal rational validation process and confirmed by the studies of the programme in practice, Chapter 4 and 5. The lower rating of the assessment on organisational skills and professionalism was probably not a disagreement on the content, as the respondents rated the other assessment on similar aspects high. We conclude that the outline of the ITA programme has sufficient content validity defined as the clinical experts' agreement on the composite content and on the pre-defined standards for making decisions about the trainees.

Research question 3: Do clinicians prefer task specific checklists or global scoring forms with categories of competence when assessing trainees' performance? Is clinicians' agreement on pass/fail decisions better when using checklists compared to global scoring forms?

The internal rational validation process resulted in choosing task specific checklists over global scoring forms addressing general aspects of competence based on the assumption that checklists might serve as a help to the assessors and trainees in clarifying expectations of performance and checklists would provide better accuracy in scoring trainees. Research question 3 was studied using an experimental design with four simulated clinical assessment scenarios and a representative group of 32 clinicians randomised into using either of two scoring formats for assessing a resident's performance. One group used a checklist with task specific items and the other a global rating form with general dimensions of competence. Clinicians rated the appropriateness of the checklists significantly higher than the global scoring format. The studies on experience with the programme in practice presented in Chapter 4 and 5 confirmed the appreciation of checklists by trainers as well as trainees. The results in chapter 3 also demonstrated a poor agreement on pass/fail decisions irrespective of scoring form used. That was explained by assessor leniency rather than lack of vigilance in the observations or disagreements over standards of performance. Assessor leniency and non-compliance with the assessment protocol were confirmed in the studies on the implementation of the programme and how the programme works in practice, Chapters 4 and 5. We conclude that clinicians prefer task specific checklists to

global scoring forms in a postgraduate clinical context. However, checklists are no guarantee of consistency in clinicians' scoring of trainees and the internal rational validation process did not foresee the problems with leniency and non-compliance with the protocol.

Research question 4: To what extent can ITA be implemented? How well is ITA accepted by the trainees? What are the positive and negative aspects?
No matter how extensive the internal rational validation process and the prospective consensus seeking, barriers to implementing new concepts and imposing ITA in a busy work-based environment can be expected. Research question 4 was studied by of a nation-wide survey addressing the issues of implementation and preliminary evaluation data on trainees' opinion about the ITA programme. Reponses were obtained from 25 (96%) programme directors who reported that in total 83 (83%) trainees had been enrolled in the ITA programme. These results were later confirmed in the study reported in Chapter 6, where between 77% and 85% of the trainees reported being enrolled in the programme. In general the programme was well accepted by both trainees and trainers. Barriers to implementation were time constraints and resistance among some of the senior colleagues. The summative parts of the programme were more frequently applied than the formative parts. The positive aspects of the programme included making goals and objectives clear, being of help in structuring training and learning, and in documenting and monitoring trainees' progress. We conclude that the ITA programme was widely implemented and accepted by trainees 1½ year after the introduction of the programme. The positive aspects were the advantage of structure and the negative aspects were mainly time constraints.

Research question 5: What is the educational impact of the programme in practice? How does the programme work in actual practice and how does that affect the training, teaching and learning?
The possible benefits and the educational impact of ITA are described in the literature. However, the literature on how ITA works in actual practice is sparse. Research question 5 was studied in three clinical departments through semi-structured interviews with programme directors, clinical supervisors, and trainees. In summary the perceived benefits of the programme were structuring training, teaching, and learning, drawing attention to other aspects of competence than exclusively medical expertise, fostering learning and motivating all parties to study more, and being of help in managing problem residents. Issues of uncertainties about standards and conflict with service demands declined over time and experience with the programme and the departments worked towards a more structured planning of the assessments. The results dem-

onstrated some variance in how the ITA programme was administered in practice and that affected the perceived value of ITA. Timing and sequencing of the programme were important. The users expressed that in order to be meaningful assessments should be placed at an appropriately early stage of training and used as a licence to practice un-supervised rather than after the trainees had been practicing independently for some time. Secondly, the trainers' attitude and the rigour in assessment practice had an impact on the trainees' perception of learning from the assessment and the meaningfulness of doing assessments. Finally, the educators and learners emphasised that the assessments should include a challenge and a learning point rather than being mere documentation of competence.

In conclusion the results of this study support the validity of the design of the ITA programme. Timely and rigorous assessment practice influenced the perceived benefit and educational impact of ITA. The results indicate that ITA might be of help in introducing new and broader concepts of competence in the medical profession and the study gives a first impression of how the organisation learns through the experience with ITA.

Research question 6: What is the effect of ITA in postgraduate education on the trainees' clinical confidence?

It was intuitively hypothesised that ITA drives learning and an increase in trainees' confidence in clinical practice was expected. We studied the last research question of this thesis by surveying junior doctors clinical confidence before (year 2001) and two years after (year 2003) the introduction of ITA. Responses to a questionnaire including 155 items on broad aspects of clinical competence were acquired from 377 (71%) doctors in 2001 and from 344 (66%) in 2003. Junior doctors represented house staff, introduction-year trainees, residents, senior residents, and specialists. Clinical confidence did not increase after the introduction of ITA, neither in aspects related to the ITA programme or in other aspects. This is contradictory to the results presented in Chapter 4 and 5, where the user of the programme indicated that ITA fostered learning. It is possible that the perceived confidence in clinical practice is not related to competence in terms of actual knowledge and skills. As discussed in Chapters 5 and 6, lack of timely application of assessments according to the trainees' individual practice might in part explain why ITA did not have an impact on trainees' confidence. We conclude that ITA did not have an effect on trainees' clinical confidence.

Conclusion

In summary theories on education and previous research outcomes informed the design and content of a specific ITA programme for 1st year trainees in anaesthesiology. The design included a mastery oriented, sequenced programme emphasising skills in the beginning of the training year and reflection on practice at later stages of the year. Based on the theoretical considerations checklists were preferred over global rating scales as instruments for the skills assessments and these checklists included theoretical questions on the individual skills assessed. Validity of this design was confirmed by clinical experts' agreement on the content and appreciation of checklists over global scoring forms. The programme proved feasibility and was widely implemented rather shortly after the introduction. Both trainers and trainees appreciated the programme for being of help in structuring teaching, training, and learning, and the users expressed a perceived benefit of ITA in fostering learning. However, that was not reflected in trainees' level of clinical confidence 2 years after introduction of ITA. Studies on the programme in practice demonstrated some variance in compliance with the assessment protocol. The value of ITA seemed to be linked to how the programme was administered in practice and timely, sequenced assessment according to trainees' professional development was crucial.

Discussion of methodological aspects

In this section the methodological aspects of the composite studies in this thesis are discussed under the headings: psychometric considerations, feasibility and educational impact.

Psychometric considerations

Psychometric properties of an assessment instrument can only be judged in terms of data collected from authentic application.[1] The six research questions in this thesis relate to the construction and introduction of an ITA programme and the first experience with the programme. The study of validity in Chapter 2 was confined to surveying consultants' opinions on the outline of the programme. This study is but a start and there are many other important questions related to the validity of ITA that have to be studied at a later stage.[2] Concurrent validity of new assessment strategies are often estimated by comparisons to other measures of competence such as traditional end-of training exams and OSCE formats. That was not possible in the context of our ITA programme, as formal summative assessments or in-training evaluation of trainees in postgraduate education did not exist in our country. Evidence of construct

validity was indirectly demonstrated in Chapter 6, where junior doctors' clinical confidence on aspects related to the ITA programme increased according to levels of training and experience. In the future direct evidence of construct validity could be obtained by applying the assessments to a wide range of doctors with increasing levels of expertise.[3-5] Discriminative validity as to whether the instruments can distinguish between poor and correct performance was indicated in the study presented in Chapter 3. Clinical assessors gave significantly lower scores to three simulated video taped performances, where the trainee made deliberate mistakes compared to the scenario, in which the trainees performed correctly. However, there were large variances in the ratings and pass/fail decisions were inconsistent among assessors in the first three scenarios, whereas all assessors passed the trainee in the last one. Holmboe *et al.*[6] used a similar approach in demonstrating overall discriminative validity of the mini-clinical examination assessment instrument (MiniCEX). Like in our study, they found a wide range in ratings among the assessors. Although checklists might improve accuracy in scoring,[7] they are no guarantee for examiner consistency in pass/fail decisions.[7,8]

The results in Chapter 3 indicate poor reliability of the individual assessments. In ITA sampling across many situations and use of several different assessors alleviate these problems and the composite programme might have sufficient reliability.[9-13] Studying psychometric properties on a sub-set of the entire ITA programme cannot be a substitute for studying the composite programme as it works in practice. In internal medicine high reliability has been demonstrated for both the Mini-CEX, which is based on direct observation of trainee performance, and the American Board of Internal Medicine monthly evaluation form (ABIM-MEF).[14-17] These instruments are applied on several occasions over time and performance scores on corresponding elements of these instruments are highly correlated.[14] The high reliability and concurrent validity could be a result of the assessors using mainly the upper part of the scoring scale in both instruments and as a result the instruments are poor in demonstrating trainees' increase in performance over time.[14-16] From a psychometric point of view the advantage of the high reliability appears to be a trade off regarding validity. There are several possible explanations to this scale shrinking. First, the majority of the trainees might be in the top range. In postgraduate education trainees are a highly selected group of learners and the fact that the vast majority of trainees receive scores between 7 and 9 on a 9-point scale corresponds to the prevalence of problem trainees being only around 7%.[15,18] In the study reported in Chapter 5 the programme directors indicated that ITA was of help in identifying problem trainees at an earlier stage as well as specifying remediation plans. The value of ITA is that the staff gets to observe

trainees in actual practice more than just occasionally. Most probably this serves as a screening for minimal performance standards[19] and is simultaneously used for feedback and instruction when needed.[16] As discussed in Chapter 3, clinicians' conflicting role of supporting learning and development as well as being a judge could explain the phenomenon of inconsistency in pass/fail decisions on the individual assessments. Suggestions to solve that problem include appointing clinicians to only one of the roles or by leaving the summative pass/fail decisions to a committee.[19-21] In Chapter 5 trainees indicated that some assessors are lenient and are satisfied with minimal acceptable levels of performance of their trainees. In some cases this might reflect deficiencies in their own performance standards and underlying competence.[22-26] Finally, in the context of our studies unfamiliarity with the concept of standardised assessment of clinical competence might in part explain the inconsistency in assessors' scoring. However, studies on the Mini-CEX demonstrate similar large inter-rater variability despite almost ten years experience with this instrument.[6]

Several authors emphasise the need for training the clinicians as assessors in ITA.[6,27,28] Our strategies of preparing the clinicians for the role as assessors were apparently insufficient in ensuring adherence to the protocol and rigorous assessment. However, training assessors might have only minor or no effect[7,21,29,30] and further research along that line could be a dead end. Rather future studies on what internal and external mechanisms influence assessors' decision-making and dedication would be of interest.[21] Such mechanisms might include the effect of benchmarking with other measures of outcome,[31] involving the assessors in the whole process of constructing the instruments,[8] group decisions,[21] and external review of assessment practice.[32]

Feasibility

Feasibility as to the implementation of our ITA programme was rather high compared to other composite programmes[33] and more than 80% of trainees were enrolled in the programme rather shortly after the introduction, Chapter 4 and 6. The perception of the programme being a benefit to the users in structuring teaching, training, and learning might well have contributed to the wide implementation, Chapters 4 and 5. However, the number of research studies and surveys performed in relation to the introduction of this particular ITA programme probably had an impact on the implementation as a Halo effect.[34] Introducing change by concurrent research studies is known to support implementation of innovations in medical education.[34] The validity study reported in Chapter 2 served the purpose of informing the clinical departments nation-wide about the programme and drawing attention to the elements of the pro-

gramme and standards of pass/fail decisions. The study on the simulated trainee performance afforded much discussion about standards of performance and various clinical departments subsequently used the videos in faculty development programmes. The studies on implementation and the effect on trainees' confidence contributed to a continuous follow up on this new initiative. This ITA programme was the first of its kind in the country and that could have afforded a pioneer spirit. All of these aspects are included in recommendations for managing change in medical education.[34]

Implementation of the programme was studied through questionnaire surveys. Response rates were generally high and we feel confident in concluding on the results. However, a more substantiate evidence of the implementation would be a continuous collection of documentation of assessments performed. That was not possible in the context of this specific ITA programme, as no authorities requested systematic reporting on educational results from graduate medical programmes. However, keeping national databases on assessment initiatives is highly recommendable for the monitoring and improving the quality of the programme over time and benchmarking across programmes or jurisdictions.[35-37] Evaluation of data collected over a time is necessary for research into psychometric properties and databases are valuable in studying the effect of new initiatives and reviewing existing educational practices accordingly.[2,15,16,32,38,39] Coupling educational databases with the growing number of clinical databases holds promises for future studies on the predictive validity of assessment practices and programme characteristics for clinical outcome and quality of clinical practice.[40-42]

Educational impact

The studies on the educational impact included a qualitative study, Chapter 5, and a study of trainees' clinical confidence before and after the introduction of ITA, Chapter 6. As discussed in these chapters, further research is needed in order to measure the effect of ITA on actual competence in terms of knowledge and skills. One way to pursue this question in the future could be elaborating a written test on cognitive knowledge and knowledge of skills. For research purposes, measures like that have been demonstrated to be valid estimates in comparing programmes and studying the effect of educational innovations on cohorts of trainees.[43-46] Although knowledge of skills tests cannot be used for assessing individuals' actual performance it might be a valuable supplement to ITA, especially if applied as a progress test.[45] Secondly, full-scale simulation could be used to study some of the other aspects of competence such as clinical decision-making, crises-resource management, interpersonal skills, teamwork, and professionalism.[47-49] However, despite simulation

technology especially in anaesthesiology is rather advanced only few studies have investigated the psychometric properties and efficiency of these methodologies.[48]

Summary
The six research questions in this thesis relate to the construction and introduction of an ITA programme and the first experience with the programme. The strength of the studies lies in the process and considerations in outlining the programme and the developmental approach in the preliminary studies on the programme in practice. A number of future studies are required to estimate the psychometric properties of ITA and the impact on learning in terms of gain in actual knowledge and skills. Keeping national databases on assessments performed and linking those to other measures of competence and indicators of clinical practice is recommended.

Implications

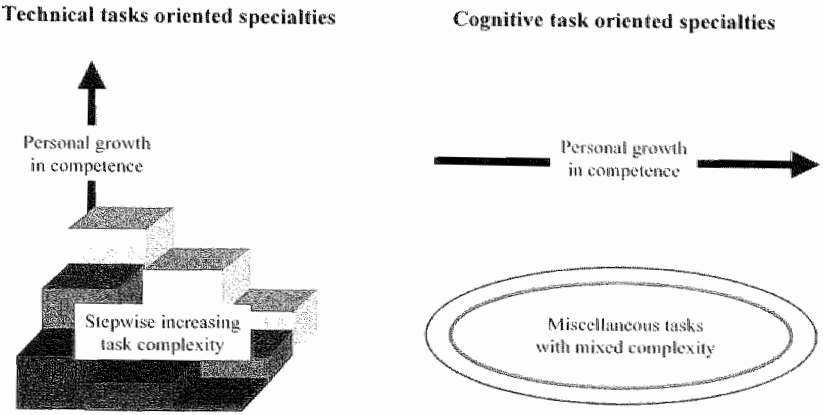
Our programme was designed for 1st year trainees in anaesthesiology and that has several implications as to the generalisability of the specific model. This section is a reflection on ITA in a wider perspective and a discussion of implications to other specialties and higher levels of training. The aspects will be addressed according to the medical and the non-medical aspects of competence.

Medical aspects of competence
The various medical specialties differ substantially in their work place organisation and their approach to professional development of their trainees. An ITA strategy suitable for one specialty might not be appropriate for another. The most striking difference among specialties is whether they are predominantly 'technical task oriented' or 'cognitive task oriented' specialties. This classification is far too simple, but is used here as an example of how this might influence the appropriateness of various ITA strategies in postgraduate education, Figure 1.

In 'technical task oriented' specialties like anaesthesiology and surgery the main training strategy is assigning the trainee increasingly complex tasks over the years, starting with rather simple tasks, standard cases, and one task at time and gradually presenting the trainees to more complicated tasks and patients in increasingly complex environments.[50] The personal growth in competence is like a stepwise incremental bricklaying construction and is ruled by skills acquisition through practicing and experience.[22] The clinical teaching strategies include instruction, di-

rect supervision, and scaffolding. In these specialties mastery oriented, sequenced task specific ITA strategies like the one presented in this thesis and the one presented by Long in neurosurgery seem appropriate.[51]

Figure 1. Personal growth in trainees' competence in 'Technical task oriented' specialties, ex. anaesthesiology and surgery and 'Cognitive task oriented' specialities, ex. internal medicine and family medicine.



In 'cognitive task oriented' specialties like internal medicine, family medicine and paediatrics the complexity of task and context is not used strategically in the same way. Trainees are assigned to a rather unselected miscellaneous group of tasks and patients in whatever context they work such as ward rounds, admittance, and ambulatories.[16] Although task complexity varies across settings and patients, mean task complexity does not vary across time in the training year.[16] The main strategy here is to let the trainee take care of increasingly wider aspects of the individual tasks.[39,52] The clinical teaching strategies include more distant supervision, post hoc discussion of cases, and conferences.[18] The personal growth in competence is pictured as rings in the water and is ruled by knowledge acquisition through experience.[53-58] In these specialties instruments that adapts to a broad range of clinical situations and settings and include the possibility of grading performance on general characteristics of the tasks seems more feasible than mastery oriented task specific instruments.[14,17,52]

The specialty characteristics are reflected in the link between various aspects of competence and that has implications for the validity of various assessment strategies. In 'cognitive task oriented' specialties the link between knowledge and clinical performance is generally strong.[15,42] For

example in paediatrics a strong correlation (r=0.71) was found between in-training examination scores and performance in an OSCE containing predominantly interview, chart review, and interpretation stations and only a few physical examination stations.[59] As cognitive tasks are dominant in these specialties, in-training theoretical examinations can be used in monitoring trainees' progress and identifying problem trainees.[18] In contrast the relation between knowledge and clinical performance in 'technical task oriented' specialties is generally poor.[60,61] In surgery Scott *et al.*[60] demonstrated that in-training examination scores were not correlated to scores on a skills test, r=0.151, or scores on intra-operative performance, r=0.106. The acquisition of skills often comes before and eventually even without the full theoretical knowledge.[62,63] For example in anaesthesiology Sivarajan *et al.*[63] found that scores on a spinal anaesthesia skills test were significantly higher than scores on a complementary knowledge test of that skill and the two sets of scores were not correlated, r=0.19. In these specialties performance on in-training examinations and performance in practice is not linked, but rather a co-existing phenomenon[60,61] and cognitive tests are not useful in identifying problems in practice.[64]

The above-mentioned classification of specialties is much too simple and in every specialty there is a need for learning some technical as well as some cognitive oriented skills.[22,65] In Chapter 1 we discussed how different specialties have different values, traits and preferred learning styles. The various specialties probably attract trainees with certain personality characteristics[66] and the trainees are formed by the culture in which they are trained.[67] This is reflected in performance on various aspects of competence. In 'cognitive task oriented' specialties doctors are likely to have deep intellectual curiosity and focus on thinking rather than acting[67,68] and scores on cognitive tasks tend to be higher than scores on technical oriented tasks.[59,69] For example in internal medicine residents scored significantly higher on test interpretation stations than physical diagnosis stations in an OSCE[69] and mean scores on physical examination or procedural skills on the Mini-CEX and the ABIM are lower than scores on cognitive aspects of competence.[14,16] In 'technical oriented specialties' focus is on acting[67] and scores on technical performance can be significantly higher than scores on related cognitive tests.[4,61,63] For example Sloan *et al.*[4,61] found generally higher sub-scores on procedural and technical stations in an OSCE compared to interpretation stations.

These aspects have implications for choosing assessment strategies and ITA can be used strategically in driving learning in the non-dominant areas of a specialty.[10,67] In the context of our programme the non-dominant area was the cognitive one and we deliberately targeted

cognitive aspects by including theoretical questions in the checklists on procedural skills and by including reflective assignments in the ITA programme, Chapter 1. As reported in Chapter 4 and 5, the trainees appreciated the cognitive aspects and both trainees and trainers indicated that the ITA programme made them study more. Future studies are needed to demonstrate whether ITA actually has an effect on those aspects of learning and how that might influence clinical performance and quality of practice.[70,71] However, studies on the programme in practice also demonstrated that implementing elements targeting the non-dominant areas, in this case written assignments on academic aspects of competence, might meet resistance from both trainees and trainers, Chapter 2, 4, and 5. In 'cognitive task oriented' specialties the non-dominant areas are skills and a structured approach to the tasks. This means that less attention is paid to train and practice these aspects in order to become an expert.[22] For example performing ward rounds is a substantiate part of internists' practice. However, standards of efficient or 'good' ward rounds are rarely defined, taught or assessed.[5,72] A structured approach to tasks might free up mental resources to deal with the cognitive challenges[22] and increase efficiency and outcome in patient management.[72,73] Thus, 'cognitive task oriented' specialties might benefit from using ITA in driving learning of basic skills and taking a structured approach to the tasks.

In both types of specialties the need for better training and assessment of even very basic procedural skills has repeatedly been demonstrated.[69,74-76] In a postgraduate work-based context the focus in the early training is on learning a number of new procedural skills and the learning of cognitive oriented aspects of competence comes second and slower.[22,62,77] This order is opposite to the traditional undergraduate medical curriculum, but imperative for the trainees to be able to function in the postgraduate work-based context, Chapter 1. During professional development, integration of skills, knowledge, and behaviours are increasingly expected as are the ability to handle uncertainties and more complex situations.[78] In Chapter 6 we demonstrated that trainees' level of clinical confidence was related to time spent in training. Although practice and clinical experience may lead to increased confidence this is no guarantee of actual competence and quality of practice.[22,79] Quality of performance in practice depends on time spent in practice as well as the ability to assess the quality of one's own practice and an attitude towards continual improvement.[22] In postgraduate education the challenge is to support the development of an integrated, deep, holistic learning and an attitude towards self-directed learning through continuously monitoring and reflecting on quality of practice.[22,39,80-83] Future research should investigate how ITA programmes can support the development of such skills and attitude and how they should be assessed.[84]

Non-medical aspects of competence

The non-medical aspects of competence are embraced in many of the roles of future specialists such as communicator, collaborator, scholar, and professional.[85,86] The educational authorities have defined the future roles and the underlying aspects of competence as they apply to all specialists in general.[85,86] A number of new tools have been suggested for assessing these general aspects.[39,86-88] However, just as with the medical aspects of competence ITA programmes will have to address the non-medical aspects relative to characteristics of each specialty.[89] The values and characteristics of a specialty is often part of a hidden curriculum and using professionals' verbalisation of professional practice could be of help in defining appropriate objectives and constructing ITA programmes accordingly.[90-92] In addition, experience from other organisations may be valuable in defining the non-medical aspects of competence such as teamwork.[47,49,93] Future studies are needed on professional practice in order to better understand and define these aspects of competence [49,93] and qualitative research approaches in a joint collaboration between medical professionals and professional educators such as psychologists or sociologists will probably add significantly to the insight and verbalisation of these tacit areas.[70,71,90,92] Emphasising broader aspects of competence in medical education including non-medical aspects was based on societies' needs and critique of prevailing practice[94] and the present staff of clinical teachers may not be dependable in teaching and assessing professional values, attitudes, and behaviours.[23-26] Training the trainers in these aspects is necessary[95] – and specifying appropriate behaviour in ITA instruments could be of help in qualifying all parties as indicated in Chapter 5. However, much more research is needed into which strategies are effective and efficient in qualifying all parties and ensuring an outcome that meets various stakeholders' expectations.[26]

A major challenge to the professions and medical educators is defining and specifying the developmental stages that the trainees run through over the time of professional growth.[22,39,91] Various aspects of competence as they apply to the future roles of specialists are emphasised at various levels of training and professional development.[22,39,82] For example issues of professionalism relevant to the specialist such as the social contract, codes of ethics, participation in professional societies, and altruism are not on the agenda in the trainees' daily activities.[96] Experts in a domain often have difficulties in verbalising both medical and non-medical aspects of competence[78] and they may not be aware of the problems trainees experience related to their specific tasks and contexts. In the process of defining objectives and tailoring programmes and ITA instruments to the various training levels it may be helpful to include trainees and other stakeholders working closely with the trainees.[97,98]

Summary
Medical specialties differ substantially in preferred training strategies and culture according to whether they are primarily 'technical task oriented' or 'cognitive task oriented' specialties. This has implications for choice of ITA strategies and a model suitable for one specialty might not be appropriate in another. ITA programmes should be tailored to the trainees' professional development and relevant daily work activities and support the acquisition of basic clinical skills as well as deep, holistic aspects of competence. ITA can be used to drive learning in non-dominant areas of a specialty.

Conclusion and future research

The overall research question of this thesis was. "How do theory on education and previous research outcomes inform the design and content of ITA in a work-based postgraduate medical education context and how does that apply to the feasibility and educational impact of ITA in actual practice? This research question was studied through generation of six sub-questions applied to a context of a specific ITA programme designed for 1st year trainees in anaesthesiology and a subsequent theoretical discussion as to the generalisation of the specific model and implications of the findings in a wider perspective.

Conclusion on the first part of the research question is that theory on education and previous research outcomes inform that ITA strategies appropriate in one specialty might not be appropriate in another. Design and content of composite ITA programmes must address trainees' levels of experience and work-related context and support the development of broad, holistic aspects of competence. In technical task oriented specialties mastery oriented, sequenced programmes seem more appropriate than in 'cognitive task oriented' specialties, where global instruments adapting to a wide variety of clinical situations and settings could be a better choice. Irrespective of specialty, ITA can be used in driving learning in the non-dominant areas. In 'technical task oriented' specialties this includes the cognitive aspects of competence and in 'cognitive task oriented' specialties this includes focus on basic skills and structure in the approach to the tasks. In all of the medical specialties there is a need for defining standards of performance in both medical and non-medical aspects of competence. Using checklists and sufficiently elaborate instruments can be of help in making standards clear to trainees and assessors.

The conclusion on the last part of the research question is drawn from applying the theories to a specific ITA programme designed for 1st year

trainees in anaesthesiology and studies of this programme in actual practice. The programme proved feasible as to the implementation of ITA. Studies on the programme in practice demonstrated some variance in compliance with the assessment protocol. Nevertheless, the users found the programme helpful in guiding teaching, training, and learning and in identifying and managing problem trainees. Trainees and clinical teachers appreciated the details in the checklists and clinical assessors found checklist more appropriate than global rating forms. The users of the programme indicated that the cognitive aspects of the programme made all parties study more. The programme had no effect on level of trainees' clinical confidence measured before and two years after introduction of ITA. The perceived value of ITA seemed to be linked to how the programme was administered in practice and timely, rigorous assessment was crucial. There are reasons to believe that the clinical departments learn over time and experience with the programme and the full potential of the specific programme remains to be explored. The impact of ITA on actual competence in terms of knowledge and skills remains to be studied in the future.

The process and considerations involved in outlining the specific ITA programme presented in this thesis might serve as a template for those who wish to design similar programmes. The specific model will have to be adapted to the characteristics of the specialty and training level. The studies presented in this thesis are but a start and points to several aspects of ITA to be studied in the future:

- Are specialty specific ITA models better than global models assessing general aspects of competence defined as an integrative habit of action?
- Are ITA models tailored to trainees' professional development more efficient than global models addressing general aspects of competence defined as an integrated habit of action?
- How important is challenging the non-dominant areas of the specialties and will that contribute to an effect on actual competence and quality of practice?
- Does ITA on clinical performance alone suffice in ensuring trainees' achievement of competence or is assessment of actual skills and knowledge a necessary compliment as to ensure quality of performance in practice?
- What internal and external mechanisms influence clinicians' assessment practice? What is the effect of benchmarking with other measures of outcome and external review of assessment practice? How can clinician assessors effectively be prepared for ITA?

- What is the effect of various models of ITA on quality of clinical practice in training as well as in subsequent continuous practice?
- What are the psychometric properties of composite ITA programmes addressing broad aspects of medical and non-medical aspects of competence and what kind of evidence is useful for estimating that?
- How can ITA foster development of an integrated, deep, holistic learning and an attitude towards self-directed learning and an attitude towards continuously monitoring and reflecting on quality of practice? How can these aspects be assessed?
- How can national databases on assessment, educational practice and clinical practice be established and used without fostering a repressive culture of fear, shame and blame?

Constructing meaningful ITA programmes as they apply to the work-based postgraduate education context in a wide variety of specialties and cultures is a huge challenge to educators. The approach taken to develop the specific ITA programme and studying the programme in practice is recommended. Establishing national databases on trainees and assessment practice across programmes would be of help in the studies of feasibility and educational impact of various strategies. ITA seems to be a valuable strategy in addressing broad aspects of clinical performance and integrating assessment in the learning process. However, due to problems of assessor bias in clinical practice ITA may not be sufficient in providing evidence of expected learning outcomes and supplementary assessment strategies might be needed.

References

1. Dauphinee WD, Blackmore DE. Assessing the assessors' assessment. Med Educ 2001; 35: 317-8.
2. Hutchinson L, Aitken P, Hayes T. Are medical postgraduate certification processes valid? A systematic review of the published evidence. Med Educ 2002; 36: 73-91.
3. Hodges B, Regehr G, Hanson M, McNaughton N. Validation of an objective structured clinical examination in psychiatry. Acad Med 1998; 73: 910-2.
4. Sloan DA, Donnelly MB, Schwartz RW, Felts JL, Blue AV, Strodel WE. The use of the objective structured clinical examination (OSCE) for evaluation and instruction in graduate medical education. J Surg Res 1996; 63: 225-230.
5. Norgaard K, Ringsted C, Dolmans D. Validation of a checklist to assess ward round performance in internal medicine. Medical Education 2004 (in press).
6. Holmboe ES, Huot S, Chung J, Norcini J, Hawkins RE. Construct validity of the Mini-Clinical evaluation exercise (MiniCEX). Acad Med 2003; 78: 826-830.
7. Noel GL, Herbers JE, Caplow MP, Cooper GS, Pangaro JN, Harvey J. How well do internal medicine faculty members evaluate the clinical skills of residents? Ann Int Med 1992; 117: 757-65.

8. Wilkinson TJ, Frampton CM, Thompson-Fawcett M, Egan T. Objectivity in objective structured clinical examinations: checklists are no substitute for examiner commitment. Acad Med 2003; 78: 219-23.
9. Hays RB, Fabb WE, van der Vleuten CPM. Reliability of the fellowship examination of the Royal Australian College of General Practitioners. Teach Learn Med 1995; 7: 43-50.
10. Van der Vleuten CPM. The assessment of professional competence: developments, research and practical implications. Adv Health Sci Educ, 1996: 1: 41-67.
11. Replogle WH. Interpretation of the American board of family practice in-training examination. Fam Med 2001; 33: 98-103.
12. Wass V, McGibbon D, van der Vleuten C. Composite undergraduate clinical examinations: how should the components be combined to maximized reliability? Med Educ 2001; 35: 325-330.
13. Crossley J, Davies H, Humphris G, Jolly B. Generalisability: a key to unlock professional assessment. Med Educ 2002; 36: 972-978.
14. Durning SJ, Cation LJ, Jackson JL. The reliability and validity of the American Board of Internal Medicine monthly evaluation form. Acad Med 2003; 78: 1175-82.
15. Durning SJ, Lannie JC, Markert RJ, Pangaro LN. Assessing the reliability and validity of the mini-clinical evaluation exercise for internal medicine residency training. Acad Med 2002; 77: 900-4.
16. Norcini JJ, Blank LL, Duffy FD, Fortna GS. The Mini-CEX: A method for assessing clinical skills. Ann Intern Med 2003; 138: 476-81.
17. Norcini JJ, Blank LL, Arnold GK, Kimball HR. Examiner differences in the Mini-Cex. Adv Health Sci Educ 1997; 2: 27-33.
18. Yao DC, Wright SM. National survey of internal medicine residency programme directors regarding problem residents. JAMA 2000; 284: 1099-1104.
19. Gordon MJ. Cutting the Gordian Knot: A two-part approach to the evaluation and professional development of residents. Acad Med 1997; 72: 876-880.
20. Yudkowsky R. Can resident evaluations demonstrate increases in residents' skills over time? Acad Med 1999; 74: S108-10.
21. Williams RG, Klamen DA, McGaghie WC. Cognitive, social and environmental sources of bias in clinical performance ratings. Teach Learn Med 2003; 15: 270-92.
22. Guest CB, Regehr G, Tiberius RG. The life long challenge of expertise. Med Educ 2001; 35: 78-81.
23. Paice E, Heard S, Moss F. How important are role models in making good doctors? BMJ 2002; 325: 707-10.
24. Vernooij-Dassen MJ, Ram PM, Brenninkmeijer WJ, Franssen LJ, Bottema BJ, van der Vleuten CP, Grol RP. Quality assessment in general practice trainers. Med Educ 2000; 34: 1001-6.
25. Iirola T, Lund VE, Katila AJ, Mattila-Vuori A, Pälve H. Teaching hospital physicians' skills and knowledge of resuscitation algorithms are deficient. Acta Anaesthesiol Scand 2002; 46: 1150-4.
26. Kenny NP, Mann KV, MacLeod H. Role modelling in physicians' professional formation: reconsidering an essential but untapped educational strategy. Acad Med 2003; 78: 1203-10.
27. Turnbull J, Gray J, MacFadyen. Improving in-training evaluation programs. JGIM 1998; 13: 317-23.
28. Feletti G, Cameron D, Dawson-Saunders B, Des Groseilliers J, Dooley B, Farmer E, McAvoy P. In-training assessment. In: Newble D, Jolly B, Wakeford R, eds. The certification and recertification of doctors: Issues in the assessment of clinical competence. Cambridge University Press. 1994: 151-66.
29. Haber RJ, Avins AL. Do ratings on the American Board of Internal Medicine resident evaluation form detect differences in clinical competence? J Gen Intern Med 1994; 9: 140-5.

30. Newble DI, Hoare J, Sheldrake PF. The selection and training of examiners for clinical examinations. Med Educ 1980; 14: 345-349.
31. MacRae HM, Vu NV, Graham B, Word-Sims M, Colliver JA, Robbs RS. Comparing checklists and databases with physicians' ratings as measures of students' history and physical examination skills. Acad Med 1995; 70: 313-7.
32. Fowell SL, Southgate LJ, Bligh JG. Evaluating assessment: the missing link? Med Educ 1999; 33: 276-81.
33. Houghton G, Wall D. Trainers' evaluations of the West Midlands formative assessment package for GP registrar assessment. Med Teach 2000; 22: 399-405.
34. Gale R, Grant J. AMEE Medical education guide No. 10: managing change in a medical context: guidelines for action. Med Teach 1997; 19: 239-49.
35. Garibaldi RA, Subhiyah R, Moore ME, Waxman H. The in-training examination in internal medicine: an analysis of residents' performance over time. Ann Intern Med 2002; 137: 505-10.
36. Hall JR, Cotsonis GA. Analysis of residents' performances on the in-training examination of the Americal Board of Anesthesiology- American Society of Anesthesiologists. Acad Med 1990; 65: 475-7.
37. Wolfshal SD, Beasley BW, Kopelman R, Stickley W, Gabryel T, Kahn MJ, Membership survey and scientific data committee, the association of program directors of internal medicine. Benchmarks of support in internal medicine residency training programs. Acad Med 2002; 77: 50-6.
38. Paice E, Aitken M, Cowan G, Heard S. Trainee satisfaction before and after the Calman reforms of specialist training: questionnaire survey. BMJ 2000; 320: 832-6.
39. Epstein RM, Hundert EM. Defining and assessing professional competence. JAMA 2002; 287: 226-35.
40. Sharp LK, Bashook PG, Lipsky MS, Horowitz SD, Miller SH. Specialty board certification and clinical outcomes: The missing link. Acad Med 2002; 77: 534-42.
41. Whitcomb ME. It's time to focus on the quality of GME. Acad Med 2003; 78: 1-2.
42. Tamblyn R, Abrahamowicz M, Dauphinee WD, Hanley JA, Norcini J, Girard N, Grand'Maison P, Brailowsky C. Association between licensure examination scores and practice in primary care. JAMA 2002; 288: 3019-26.
43. Remmen R, Scherpbier A, Denekens J, Derese A, Hermann I, Hoogenboom R, Van der Vleuten C, Van Royen P, Bossaert L. Correlation of a written test of skills and a performance based test: a study in two traditional medical schools. Med Teach 2001; 23: 29-32.
44. Ram P, Van der Vleuten C, Rethans J-J, Schouten B, Hobma S, Grol R. Assessment in general practice: the predictive value of written-knowledge tests and a multiple-station examination for actual medical performance in daily practice. Med Educ 1999; 33: 197-203.
45. Kramer AW, Jansen JJ, Zuithoff P, Dusman H, Tan LH, Grol RP, van der Vleuten CP. Predictive validity of a written knowledge test of skills for an OSCE in postgraduate training for general practice. Med Educ 2002; 36: 812-9.
46. Van der Vleuten CPM, Van Luyk SJ, Beckers HJM. A written test as an alternative to performance testing. Med Educ 1989; 23: 97-107.
47. Glavin RJ, Maran NJ. Integrating human factors into the medical curriculum. Med Educ 2003; 37(Suppl): 59-64.
48. Byrne AJ, Greaves JD. Assessment instruments used during anaesthetic simulation: review of published studies. Br J Anaesth 2001; 86: 445-50.
49. Fletcher GCL, McGeorge P, Flin RH, Glavin RJ, Maran NJ. The role of non-technical skills in anaesthesia: a review of current literature. Br J Anaesth 2002;88:418-29.
50. Patel BS, Feerick A. Will competency assessment improve the training and skills of the trainee anaesthetist? Anaesthesia 2002;57:711-2.
51. Long DM. Competency-based residency training: the next advance in graduate medical education. Acad Med 2000; 75: 1178-83.

52. Pangaro L. A new vocabulary and other innovations for improving descriptive in-training evaluations. Acad Med 1999; 74: 1203-7.
53. Steel K, Norcini J, Brummel-Smith K, Erwin D, Markson L. The first certifying examination in geriatric medicine. J Am Geriatr Soc 1989; 37: 1188-91.
54. Norcini JJ, Grosso LJ, Shea JA, Webster GD. The relationship between features of residency training and ABIM certifying examination performance. J Gen Intern Med 1987; 2: 330-6.
55. Norcini JJ, Lipner RS. The relationship between the nature of practice and performance on a cognitive examination. Acad Med 2000; 75(Suppl): S68-70.
56. Norcini JJ, Shea JA, Langdon LO, Hudson LD. First American Board of Internal Medicine critical care examination: process and results. Crit Care Med 1989; 17: 695-8.
57. Norcini JJ. Indicators of educational effectiveness of subspecialty training programs in internal medicine. Acad Med 1995; 70: 512-6.
58. Kramer AWM, Düsman H, Tan LHC, Jansen KJM, Grol RPTM, Van der Vleuten CPM. Effect of extension of postgraduate training in general practice on the acquisition of knowledge of trainees. Family Practice 2003; 20: 207-12.
59. Joorabchi B. Objective structured clinical examination in a pediatric residency program. AJDC 1991; 145: 757-62.
60. Scott DJ, Valentine RJ, Bergen PC, Rege RV, Laycock R, Tesfay ST, Jones DB. Evaluating surgical competency with the American Board of Surgery in-training examination, skill testing, and intraoperative assessment. Surgery 2000; 128: 613-22.
61. Sloan DA, Donnelly MB, Schwartz RW, Strodel WE. The objective structured clinical examination. The new gold standard for evaluating postgraduate clinical performance. Ann Surg 1995; 222: 735-42.
62. Rasmussen J. Skills, rules, knowledge, signals, signs, and symbols and other distinctions in human performance models. IEEE Transaction Systems, man and Cybernetics 1983; 123: 257-66.
63. Sivarajan M, Miller E, Hardy C, Herr G, Liu P, Willenkin R, Cullen B. Objective evaluation of clinical performance and correlation with knowledge. Anesth Analg 1984; 63: 603-7.
64. Cusimano MD, Yonke AM, Tucker WS. An analysis of attrition from Canadian neurosurgery residency programs. Acad Med 1999; 74: 925-31.
65. Moorthy K, Munz Y, Sarker SK, Darzi A. Objective assessment of technical skills in surgery. BMJ 2003; 327: 1032-7.
66. Hojat M, Brigham TP, Gottheil E, Xu G, Glaser K, Veloski JJ. Medical students' personal values and their career choices a quarter-century later. Psychol Rep 1998; 83: 243-8.
67. Baker JD, Cooke JE, Contoy JM, Bromley HR, Hollon MF, Alpert CC. Beyond career choice: the role of learning style analysis in residency training. Med Educ 1988; 22: 527-32.
68. Hojat M, Nasca TJ, Magee M, Feeney K, Pascual R, Urbano F, Gonnella JS. A comparison of the personality profiles of internal medicine residents, physician role models, and the general population. Acad Med 1999; 74: 1327-33.
69. Dupras DM, Li JTC. Use of an objective structured clinical examination to determine clinical competence. Acad Med 1995; 70: 1029-34.
70. Klemola U-M, Norros L. Analysis of the clinical behaviour of anaesthetists: recognition of uncertainty as a basis for practice. Med Educ 1997; 31: 449-56.
71. Klemola U-M, Norros L. Practice-based criteria for assessing anaesthetists' habits of action: outline for a reflexive turn in practice. Med Educ 2001; 35: 455-64.
72. Pronovost P, Berenholtz S, Dorman T, Lipsett PA, Simmonds T, Haraden C. Improving communication in the ICU using daily goals. J Crit Care 2003; 18: 71-6.
73. Risser D, Simon R, Rice M, Salisbury M. The potential for improved teamwork to reduce medical errors in the emergency department. Ann Emergency Med 1999; 34: 373-83.

74. Jansen JJM, Scherpbier AJJA, Metz JCM, Grol RPTM, van der Vleuten CPM, Rethans JJ. Performance-based assessment in continuing medical education for general practitioners: construct validity. Med Educ 1996; 30: 339-44.
75. Kelly MH, Campbell LM, Murray TS. Clinical skills assessment Br J Gen Prac 1999; 49: 447-50.
76. Fox RA, Clark CLI, Scotland AD, Dacre JE. A study of pre-registration house officers' clinical skills. Med Educ 2000; 34: 1007-12.
77. Marel GM, Lyon PM, Field MJ, Barnsley L, Hibbert E, Parise A. Clinical skills in early postgraduate medical trainees: patterns of acquisition of confidence and experience among junior doctors in a university teaching hospital. Med Educ 2000; 34: 1013-5.
78. Dreyfus HL, Dreyfus SE. Mind over machine: The power of human intuition and expertise in the era of the computer. Oxford, Basil Blackwell. 1986.
79. Marteau TM, Wynne G, Kaye W, Evans TR. Resuscitation: experience without feedback increases confidence but not skill. BMJ 1990; 300: 849-50.
80. Coles C. Developing professional judgment. J Cont Educ Health Prof 2002;22:3-10
81. Friedman Ben-David M. The role of assessment in expanding professional horizons. Med Teach 2000; 22: 472-6.
82. Leung W-C. Competency based medical training: review. BMJ 2002; 325: 693-6.
83. Fraser SW, Greenhalgh T. Coping with complexity: educating for capability. BMJ 2001; 323: 799-803.
84. Hojat M, Nasca TJ, Erdmann JB, Frisby AJ, Veloski JJ, Gonella JS. An operational measure of physician lifelong learning: its development, components and preliminary psychometric data. Med Teach 2003; 25: 433-7.
85. CanMEDS 2000: Extract from the CanMEDS 2000 Project Societal Needs Working Group Report. Med Teach 2000; 22: 549-54.
86. ACGME Outcome project. Accreditation Council for Graduate Medical Education Web site: http://www.acgme.org. 2000. Accessed April 2003.
87. JCHMT: Higher medical training. Generic curriculum. Joint Committee for Higher Medical Training Web site: http://www.jchmt.org.uk. 2003. Accessed April 2003.
88. Murray E, Gruppen, L, Catton P, Hays R, Woolliscroft JO. The accountability of clinical education: its definition and assessment. Med Educ 2000; 34: 871-9.
89. Hojat M, Gonnella JS, Nasca TJ, Mangione S, Vergare M, Magee M. Physician empathy: definition, components, measurement, and relationship to gender and specialty. Am J Psychiatry 2002; 159: 1563-9.
90. Coles C. A qualitative look at practice. A comment on 'Analysis of the clinical behaviour of anaesthetists'. Med Educ 1997; 31: 448.
91. Greaves JD, Grant J. Watching anaesthetists work: using the professional judgement of consultants to assess the developing clinical competence of trainees. Br J Anaesth 2000; 84: 525-33.
92. Pope C, Smith A, Goodwin D, Mort M. Passing on tacit knowledge in anaesthesia: a qualitative study. Med Educ 2003; 37: 650-5.
93. Fletcher G, Flin R, McGeorge P, Glavin R, Maran N, Patey R. Anaesthetists' non-technical skills (ANTS): evaluation of a behavioural marker system. Br J Anaesth 2003; 90: 580-8.
94. Neufeld VR, Maudsley RF, Pickering RJ, Turnbull JM, Weston W, Brown MG, Simpson JC. Educating future physicians for Ontario. Acad Med 1998; 73: 1133-48.
95. Wright SM, Carrese JA. Excellence in role modelling: insight and perspectives from the pros. CMAJ 2002; 167: 638-43.
96. Brownell AKW, Côté L. Senior residents' view on the meaning of professionalism and how they learn about it. Acad Med 2001; 76: 734-7.
97. Ringsted C, Schroeder TV, Henriksen J, Ramsing B, Lyngdorf P, Jønsson V, Scherpbier A. Medical students' experience in practical skills is far from stakeholders' expectations. Med Teach 2001; 23: 412-6.

98. Skaarup AM, Henriksen A, Ringsted C. Using SHOs' verbalisation of competencies and learning needs. Presentation at the Association for Medical Education in Europe, AMEE Conference, Lisbon, Portugal 2002.

Summary

This thesis builds around the theme in-training assessment (ITA) as it applies to a work-based postgraduate medical education context, in this case anaesthesiology. By the introduction of a broad and practice-based concept of competence traditional strategies of assessment in postgraduate education comes short. In order to ensure structured development of the trainees into competent individual practitioners there is a need for a new approach in medical education using assessment as an on-going part of the learning process, and ITA strategies seem to be a means to meet that end. The special conditions for postgraduate education being based in a busy working context and not a teaching institution offers some unique challenges to those who are responsible for constructing the programme. These challenges include compiling a programme that sufficiently covers the content and at the same time is feasible in practice without bureaucratic overloading of the users. Despite evidence that instruments using global scoring of competence is more efficient to develop and administer the instruments of the programme will need to be sufficiently specific to serve as a tool for learning and be of help to the trainers in providing feedback and monitoring progress. Finally, the programme should aim at avoiding a blame and shame culture while serving its summative purpose of ensuring performance standards and quality of care and patient safety.

The overall research question was: How do theory on education and previous research outcomes inform the design and content of ITA in a work-based postgraduate medical education context and how does that apply to a specific programme in actual practice?

Chapter one deals with the internal rational validation process and how the theory on education and previous research outcomes informed the design and content of a specific ITA programme for 1st year residents in anaesthesiology. The main results of the rational validation process were that the programme should be sequentially dispersed over the training

year with emphasis on clinical skills in the beginning of the year and other aspects of competence and reflection on practice at a later stage of the year. Further, task specific checklists were preferred over global rating forms. The programme should be narrowed down as much as possible in order to be feasible in a work-based busy clinical context. Finally tailoring content, format, and programming of ITA to level of training and to the specific specialty was important. Problems with certain aspects of competence could be expected according to which specialty the ITA programme was meant for.

The programme was elaborated by a working-group under the Danish Society of Anaesthesiology and Intensive Care Medicine. The resulting programme addressed a broad aspect of competence and included twenty-one individual elements. In the ITA programme standards of performance were defined in task specific competence cards that included a checklist and some theoretical questions applying to the procedure. Each card also included a global pass/fail decision on the specific task. Eighteen of the elements were summative and three were formative. Trainees should pass all the summative elements of the programme in order to have the entire 1st year training approved. The internal rational validation process contributed to foreseeing barriers to implementing the written assignments included in the ITA programme, which addressed academic aspects of competence and aimed at reflection on practice and developing life-long learning skills. The ITA programme and an accompanying assessment protocol were distributed to all clinical departments of anaesthesiology.

Chapter two describes the content validity of the outline of the ITA programme. This was obtained by surveying consultants' opinion of the programme in terms of the representativeness of the aspects assessed, the perceived suitability of the programme as a basis for pass/fail decisions, and the relevance and sufficiency of the content of the different assessment instruments. Two hundred and fifty one consultants (66%) responded to the survey questionnaire. More than 75% of these agreed that the ITA programme offered adequate coverage of the curriculum for first-year trainees and was appropriate for making overall pass/fail decisions. There was a sufficiently high agreement among consultants that the content of each of the 18 summative assessments were relevant and sufficient for pass/fail decisions. For two elements, one on organisational skills and one on academic aspects of competence, there was a slightly lower agreement on the importance of the element. In conclusion this study demonstrates sufficient external validity of the ITA programme. However, the study was limited to the consultants' opinion on the out-

line of the programme and aspects of validity according to actual experience with the programme need to be studied further.

Chapter three evaluated the feasibility of task specific checklists vs. global rating forms addressing general aspects of competence for the purpose of assessing trainees' clinical performance in postgraduate education. A representative sample of 32 clinicians participated in an experimental study. The participants were randomised into two groups each of which used one of the scoring formats to assess a resident's performance in four simulated clinical scenarios on videotape. The clinicians rated the appropriateness of the scoring forms on a scale 1-5. The checklist format was rated significantly higher compared with the global rating form (mean 4.6 SD 0.5 vs. mean 3.5 SD 1.4, $p<0.001$. The inter-rater agreement regarding pass/fail decisions was poor irrespective of the scoring form used. This was explained by clinicians' leniency as assessors and non-compliance with the assessment protocol rather than by lack of vigilance in the observations or disagreements on standards for good performance. Issues of reliability of the whole programme as it works in practice will have to be studied further.

Chapter four deals with the feasibility of the ITA programme and reports the results of a survey of the implementation nation-wide. Responses were obtained form 25 (96%) of the clinical departments who reported that in total 83 out of 100 possible candidates had been enrolled in the programme. The vast majority of the 21 elements of the programme were implemented although the summative elements were applied to a larger extent than the three formative elements. In general the programme was well accepted by both trainees and trainers. The positive aspects of the programme included making goals and objectives clear, being of help in structuring training, teaching, and learning, and in documenting and monitoring trainees' progress. Barriers to implementation were time constraints and resistance among some of the senior colleagues to take on the task of being assessor. The perceived benefits probably supported the implementation. However, a Halo effect of the number of research studies on this particular ITA programme might have contributed.

Chapter five reports on how the programme works in actual practice and how that affects the perceived educational impact. Information on these issues was obtained from three clinical departments by semi-structured interviews of programme directors, supervisors, and trainees. The users found the programme being of help in guiding teaching, training, and learning and in identifying and managing problem trainees. Trainees and clinical teachers appreciated the details in the checklists and indicated

that the cognitive aspects of the programme made all parties study more. Studies on the programme in practice demonstrated some variance in compliance with the assessment protocol. Three inter-related factors appeared to influence the perceived value of ITA: The link to patient safety and individual practice using assessment as a licence to practice unsupervised rather than end of training assessment. Secondly, the benefit to educators and learners in the educational process rather than mere documentation of competence, and finally the attitude and rigour of assessment practice. There are reasons to believe that the clinical departments learn over time and experience with the programme and the full potential of the specific programme remains to be explored.

Chapter six was a nation-wide study of junior doctors' clinical confidence before and two years after the introduction of ITA. The results demonstrated that confidence within a broad spectrum of clinical competence did not change as a result of introducing ITA. Although the underlying concept of confidence is complex the results indicate that ITA did not have a significant effect on clinical competence. This is contradictory to the results of the previous studies indicating that ITA fostered learning. It is likely that the perceived confidence in clinical practice is not related to competence in terms of actual knowledge and skills. Lack of timely application of assessments according to the trainees' individual practice seems to a plausible explanation of the results. Future studies on the effect of ITA on trainees' actual competence are needed.

Conclusion on the first part of the overall research question is that theory on education and previous research outcomes inform that ITA strategies appropriate in one specialty might not be appropriate in another. Medical specialties differ substantially in preferred training strategies and culture according to whether they are primarily 'technical task oriented' or 'cognitive task oriented' specialties. This has implications for choice of ITA strategies and a model suitable for one specialty might not be appropriate in another. Design and content of composite ITA programmes must address trainees' levels of experience and work-related context and support the development of broad, holistic aspects of competence. ITA can be used strategically to drive learning in non-dominant areas of a specialty.

The conclusion on the last part of the overall research question is drawn on behalf of applying the theories to the design of a specific ITA programme for 1st year trainees in anaesthesiology and studies of this programme in actual practice. The resulting programme was a sequenced, mastery oriented ITA design using checklists to specify performance expectations. The programme emphasised skills at the early stages of the training year and reflection on practise at later stages. The

programme proved feasible as to the implementation and the users appreciated the structure in the programme and the checklists. Studies on the programme in practice demonstrated some variance in assessors' rigorousness and compliance with the sequencing of the assessments. However, both trainers and assessors indicated that the perceived value of ITA in a work-based context was linked to timeliness of assessments. The studies gave a first impression of how the organisation learns through the experience with ITA and departments worked towards a better planning of assessments. The programme had no effect on trainees' level of clinical confidence measured before and two years after introduction of ITA. The impact of ITA on actual competence in terms of knowledge and skills remains to be studied in the future.

Samenvatting

Het onderwerp van dit proefschrift is 'in-training assessment' (ITA) in de specialistenopleiding, in het bijzonder de opleiding anesthesiologie. Door invoering van een brede, aan de praktijk ontleende definitie van competentie waren de beoordelingsmethoden die traditioneel in de specialistenopleiding gebruikt werden niet meer toereikend. Om een goede structuur te ontwerpen voor de ontwikkeling van assistenten-in-opleiding tot competente specialisten was een nieuwe benadering van de opleiding nodig waarbij toetsing geïntegreerd werd in het leerproces. ITA lijkt hiervoor een goede methode. Doordat de specialistenopleiding zich afspeelt in een drukke werkomgeving en niet in een opleidingsinstituut biedt dit onderwijs unieke uitdagingen aan programmaontwikkelaars. Een van deze uitdagingen is om een programma te ontwikkelen dat zowel aan inhoudelijke opleidingseisen voldoet als praktisch uitvoerbaar is en geen onnodige administratieve rompslomp met zich meebrengt voor de betrokkenen. Ondanks aanwijzingen dat het efficiënter is om instrumenten te ontwikkelen en toe te passen die algemene competentie toetsen, zal het toetsprogramma voldoende specifiek moeten zijn om niet alleen het leren van assistenten te sturen, maar ook als hulpmiddel te dienen voor docenten bij het geven van feedback en het volgen van de voortgang van de assistenten. Ten slotte dient vermeden te worden dat het programma een sfeer doet ontstaan waarin assistenten op een beschuldigende manier op hun fouten worden aangesproken, terwijl het programma wel summatieve beoordelingen mogelijk moet maken om de kwaliteit van de zorg en de veiligheid van de patiënten te kunnen waarborgen.

Centraal in dit proefschrift staat de onderzoeksvraag: Hoe kunnen onderwijskundige theorie en resultaten van onderzoek van onderwijs gebruikt worden bij het ontwerpen van vorm en inhoud van een ITA-programma in de specialistenopleiding en hoe kan een dergelijk programmaontwerp uitgewerkt worden voor de onderwijspraktijk van een specifiek opleidingsprogramma.

In *hoofdstuk 1* wordt ingegaan op de interne rationele procedure voor het valideren van toetsmethoden en de wijze waarop onderwijskundige theorie en onderzoeksresultaten vorm en inhoud van een ITA-programma voor eerstejaars assistenten in opleiding tot anesthesioloog mede hebben bepaald. De belangrijkste resultaten van de rationele procedure om de toetsing te valideren waren dat het programma gespreid diende te worden over het gehele opleidingsjaar, waarbij de nadruk in het begin van het jaar moest liggen op klinische vaardigheden, terwijl andere aspecten van competentie en reflectie op de klinische praktijk pas later in het jaar aan de orde moesten komen. Daarnaast werd de voorkeur gegeven aan specifieke checklisten boven globale scoringsformulieren. Het programma moest zo veel mogelijk tot de essenties beperkt blijven om haalbaar te kunnen zijn in een drukke klinische werkomgeving. Ten slotte was het van belang om inhoud, vorm en programmering van het ITA-programma toe te snijden op het opleidingsniveau van de assistenten en het specialisme. Er werd rekening mee gehouden dat, al naargelang het betrokken specialisme, bepaalde aspecten van competentie problemen zouden kunnen opleveren.

Het programma werd uitgewerkt door een werkgroep van de Deense Vereniging voor Anesthesiologie en Intensive Care. Het bestreek een breed competentiedomein en bestond uit 21 afzonderlijke onderdelen. De in het toetsprogramma gehanteerde beoordelingseisen waren vastgelegd op taakspecifieke competentiekaarten waarop onder andere een checklist stond en theoretische vragen over de betreffende procedure. Elke kaart beschreef ook een globale zak/slaagbeslissing voor de specifieke taak. Achttien toetsonderdelen waren summatief en drie formatief. De assistenten-in-opleiding slaagden voor het eerste jaar van de opleiding als ze een voldoende beoordeling kregen op alle summatieve onderdelen van het toetsprogramma. De interne rationele procedure om het toetsprogramma te valideren droeg bij tot vroegtijdige signalering van obstakels voor het gebruik van schriftelijke opdrachten in het ITA-programma. Deze schriftelijke toetsing betrof theoretische aspecten van competentie en was gericht op reflectie over de praktijk en het stimuleren van vaardigheden voor levenslang leren. Het toetsprogramma werd met het bijbehorende toetsprotocol verspreid onder alle klinische afdelingen anesthesiologie.

Hoofdstuk 2 beschrijft de inhoudsvaliditeit van het ontwerp voor het ITA-programma. De inhoudsvaliditeit werd bepaald aan de hand van een enquête waarin specialisten gevraagd werd hun mening te geven over de representativiteit van de onderdelen die getoetst werden, de mate waarin zij het programma geschikt achtten voor het nemen van zak/slaag-

beslissingen en de relevantie en reikwijdte van de inhoud van de verschillende toetsinstrumenten. De enquête werd ingevuld door 251 specialisten (respons 66%). Ruim 75% was van mening dat het toetsprogramma de inhoud van het eerste jaar voldoende dekte om te kunnen dienen als basis voor zak/slaagbeslissingen. Er was voldoende overeenstemming tussen de specialisten over de relevantie van de achttien summatieve toetsen en hun bruikbaarheid als basis voor zak/slaagbeslissingen. Er was wat minder overeenstemming ten aanzien van het belang van twee onderdelen, een onderdeel betreffende organisatorische vaardigheiden en een onderdeel aangaande academische aspecten van competentie. Geconcludeerd kan worden dat de externe validiteit van het ITA-programma voldoende was. Hierbij moet worden opgemerkt dat het onderzoek uitsluitend bestond uit een opiniepeiling onder specialisten over de opzet van het programma. Verder onderzoek is nodig om aspecten van de validiteit van het programma te bepalen op basis van praktijkervaringen.

Hoofdstuk 3 beschrijft een vergelijkend onderzoek naar de bruikbaarheid van taakspecifieke checklisten en globale beoordelingsformulieren voor algemene aspecten van competentie voor het ITA-programma. Er werd een experiment uitgevoerd waaraan deelgenomen werd door een representatieve steekproef van 32 specialisten. De deelnemers werden willekeurig in twee groepen ingedeeld. Elke groep beoordeelde met behulp van een van de scoringsmethoden het klinisch handelen van een assistent in vier op video opgenomen gesimuleerde klinische scenario's. De specialisten gaven hun oordeel over de geschiktheid van de beoordelingsformulieren op een vijfpuntsschaal. De checklist kreeg een significant betere beoordeling dan de globale beoordelingsformulieren (gemiddeld 4.6, SD 0.5 tegenover gemiddeld 3.5 SD 1.4, $p<0.001$). De interbeoordelaarsovereenstemming voor de zak/slaagbeslissingen was laag voor beide beoordelingsformulieren. Dit kan verklaard worden doordat de specialisten zeer mild oordeelden en zich niet hielden aan het beoordelingsprotocol en niet zozeer doordat zij onzorgvuldig observeerden of van mening verschilden over wat een goede prestatie inhield. Betrouwbaarheidsaspecten van het programma als geheel bij toepassing in de praktijk dienen nader onderzocht te worden.

Hoofdstuk 4 behandelt de haalbaarheid van het ITA-programma en beschrijft de resultaten van een enquête over de landelijke toepassing van het programma. De enquête werd ingevuld door 25 (96%) van de klinische afdelingen. Volgens de afdelingen namen 83 van de 100 potentiële kandidaten deel aan het programma. De overgrote meerderheid van de 21 programmaonderdelen werd gebruikt. De summatieve onderdelen

werden vaker gebruikt dan de drie formatieve onderdelen. Over het algemeen werd het programma door zowel opleiders als assistenten goed ontvangen. De positieve aspecten van het programma waren de duidelijk omschreven doelstellingen en eindtermen en de bijdrage die het programma leverde aan het structureren van opleiding, onderwijs en leren en de registratie en het volgen van de voortgang van de assistenten. Belemmeringen voor invoering van het programma waren tijdgebrek en bedenkingen van enkele oudere collega's ten aanzien van het optreden als beoordelaar. De waargenomen voordelen hadden naar alle waarschijnlijkheid een gunstige invloed op de invoering van het programma. Mogelijk speelde ook een rol dat er een halo-effect ontstond doordat het programma onderwerp van verschillende onderzoeken was.

In *hoofdstuk 5* wordt besproken hoe het programma in de praktijk werkte en de invloed daarvan op de ervaren onderwijskundige effecten. De gegevens voor dit onderzoek werden verkregen aan de hand van semigestructureerde interviews met opleidingsdirecteuren, begeleiders en assistenten uit drie klinische afdelingen. De gebruikers gaven aan dat het programma een positieve bijdrage leverde omdat het als leidraad diende voor onderwijs, training en leren en de opsporing en begeleiding van probleemstudenten. Zowel assistenten als opleiders waardeerden de detaillering van de checklisten. Zij gaven aan dat de cognitieve aspecten van het programma alle betrokkenen stimuleerden om meer te studeren. Onderzoeken naar de toepassing van het programma in de praktijk lieten zien dat er verschillen waren in de mate waarin men zich aan het toetsprotocol hield. Drie onderling samenhangende factoren leken van invloed te zijn op de ervaren waarde van het ITA-programma: de relatie met de veiligheid van de patiënt en de individuele praktijk, waarbij het doel van toetsing was toestemming te verlenen om zonder begeleiding te mogen werken in plaats van het geven van een oordeel aan het eind van de opleiding.

Op de tweede plaats had het programma een gunstige invloed op het onderwijsproces voor zowel opleiders als assistenten in plaats van beperkt te blijven tot het registreren van competentie. Ten slotte was er waardering voor de aandacht voor attitude en de grondigheid van de beoordeling in de praktijk. Er is aanleiding om aan te nemen dat klinische afdelingen tijd nodig hebben om ervaring op te kunnen doen met het programma en dat nog niet alle mogelijkheden van het programma verkend zijn.

Hoofdstuk 6 behelst een landelijk onderzoek naar het zelfvertrouwen van assistenten-in-opleiding betreffende hun functioneren in de klinische praktijk waarbij een vergelijking gemaakt werd tussen het zelfvertrou-

wen voor en twee jaar na invoering van het ITA-programma. De resultaten wezen uit dat het zelfvertrouwen betreffende het volledige scala aan klinische competenties niet veranderde door invoering van het toetsprogramma. Hoewel zelfvertrouwen een complex concept is, waren er geen resultaten die wezen op een significant effect van ITA op het vertrouwen in de eigen klinische competentie. Waarschijnlijk houdt het ervaren zelfvertrouwen in de klinische praktijk geen verband met competentie in termen van werkelijke kennis en vaardigheden. Het ontbreken van toetsing die direct aansluit bij actuele praktijkervaringen van de individuele assistent lijkt een voor de hand liggende verklaring voor de resultaten. Verder onderzoek is nodig om de invloed van het ITA-programma te bepalen.

De conclusie betreffende het eerste deel van de centrale onderzoeksvraag is dat onderwijskundige theorie en onderzoeksresultaten aangeven dat een ITA-programma dat geschikt is voor het ene specialisme niet automatisch geschikt hoeft te zijn voor andere specialismen. Medische specialismen vertonen aanzienlijke verschillen wat betreft voorkeuren voor opleidingsprogramma's en opleidingscultuur. Deze voorkeuren hangen samen met de verschillen tussen primair technisch of cognitief georiënteerde specialismen. Dit heeft gevolgen voor de keuze van toetsvormen. Het betekent ook dat een toetsmodel dat geschikt is voor het ene specialisme niet geschikt hoeft te zijn voor het andere. Vorm en inhoud van een ITA-programma met verschillende onderdelen moeten aansluiten bij het ervaringsniveau en de werkomgeving van de assistenten en bijdragen aan het ontwikkelen van brede en holistische aspecten van competentie. Een ITA-programma kan strategisch ingezet worden om het leren ten aanzien van minder prominente gebieden van een specialisme te bevorderen.

De conclusie ten aanzien van het laatste deel van de centrale onderzoeksvraag is gebaseerd op de toepassing van theorieën bij het ontwerpen van een ITA-programma voor eerstejaars assistenten anesthesiologie en op verschillende onderzoeken naar het functioneren van dit programma in de praktijk. Het programma bestond uit een gefaseerd toetsprogramma gericht op de beheersing van kennis en vaardigheden waarbij gebruik gemaakt werd van checklisten om de beoordelingseisen te specificeren. In het begin van het jaar lag de nadruk in het programma op vaardigheden terwijl later in het jaar de nadruk kwam te liggen op reflectie over de praktijk. Het programma bleek uitvoerbaar te zijn en de gebruikers toonden waardering voor de structuur van het programma en de checklisten. De onderzoeken naar de toepassing van het programma in de praktijk gaven enige variatie te zien wat betreft de strengheid van de beoordelaars en het aanhouden van de volgorde van de beoordelin-

gen. Niettemin gaven zowel opleiders als beoordelaars te kennen dat wat hen betrof de waarde van het toetsprogramma gelegen was in het juiste tijdstip van de beoordelingen. De onderzoeken gaven een eerste indruk van hoe een organisatie leert door ervaringen met een ITA-programma en het streven naar een betere planning van beoordelingen door de afdelingen. Een onderzoek waarin een vergelijking werd gemaakt tussen het zelfvertrouwen van de assistenten ten aanzien van hun klinisch functioneren voor en twee jaar na invoering van het ITA-programma liet geen invloed van het programma zien op dat zelfvertrouwen. De invloed van ITA op de werkelijke competentie van assistenten in termen van kennis en vaardigheden moet nader onderzocht worden.

Acknowledgements

The working group behind the ITA programme for 1st year residents in anaesthesiology were: Andersen N, Berlac PA, Bested K, Callesen T, Christensen P, Jensen E, Jensen JW, Lemholt K, Lund J, Malling B, Mandøe H, Nørregaard O, Pedersen BD, Petersen JA, Ravlo O, Ravn L, Skjelsager K, Sprehn M, Østergaard D. The time we spent in developing the ITA programme was most enjoyable and constructive. The honour of the ITA programme and its introduction nation-wide rests with this group of enthusiastic pioneers.

I deeply respect the Danish anaesthesiologists, the trainees as well as trainers, for their willingness and effort to make the ITA programme work in practice. This thesis is built around a number of studies evaluating the ITA programme. I'm indebted to all the anaesthesiologists who took their time to fill in the questionnaires and generously shared their experience with us. Their constructive comments and reflections on how the programme can be improved have been extremely valuable for the review of this programme.

The Danish Society of Anaesthesiology and Intensive Care Medicine has backed up the projects all the way and in part sponsored the implementation and evaluation of the ITA programme. They have repeatedly placed ITA on the agenda at the Society meetings over the years. Without their everlasting support this ITA programme would never have become a reality.

The Ministry of Health and The National Board of Health has in part sponsored the studies included in this thesis. A special thanks to Karsten Beck from The National Board of Health for supporting the introduction of in-training assessment in anaesthesiology at a time where this was not common practice in Denmark.

This research is a result of the visionary directors of the Copenhagen Hospital Corporation, who founded the Postgraduate Medical Institute and as such funded the studies of this thesis. I wish to thank especially director Lone de Neergaard for constantly supporting innovation in postgraduate medical education and being a promoter of the institute.

I want to express my deepest gratitude to the Danish co-authors Doris Østergaard, Lisbet Ravn, Asger Petersen, Peter Berlac, Ann-Helen Henriksen, Anne Marie Skaarup, and Jane Pallisgaard for their great efforts and contributions. Above all, I want to thank you, Doris, for the exiting and joyful collaboration we've had on these and several other projects. I cannot express how highly I appreciate that.

Hanne Panduro and Susanne Findalen have been of great help with managing the surveys and gathering the data.

I'm indebted to Deborah Davies for her editorial assistance and help with the English.

I wish to thank my former chief Svend Juul Jørgensen for encouraging me to pursue a PhD and supporting me in getting the work done.

I have had the best of supervisors for this thesis: Albert Scherpbier and Cees van der Vleuten. I'm deeply thankful for their everlasting support and encouragement. I'm impressed by their supervisor skills that reach across distances between countries. I always got immediate feedback on papers sent to them and gentle motivating mails or letters in times, when things slowed down. Their constructive criticism and professional guidance on my work has been invaluable. Thank you so much.

Curriculum vitae

Charlotte Ringsted was born in Copenhagen, Denmark on 21st September 1953. She obtained a M.D. degree in the University of Odense, Denmark, in 1978 and a certificate of Specialist in Anaesthesiology in 1991. She left clinical medicine in 1993 for a position as hospital coordinator of clinical education at Copenhagen University Hospital, Rigshospitalet. During the time at Rigshospitalet she designed and set up the first skills lab in Denmark in 1996, in collaboration with the hospital coordinator of students' clinical affairs, professor Torben Schroeder. She obtained a Master degree in Health Professions Education (MHPE) in 1997 in Maastricht University. In 1997 she was appointed the leader of a new Postgraduate Medical Institute for the Copenhagen Hospital Corporation. Her responsibilities include research and development within postgraduate medical education, curriculum design, development of assessment programmes, residency programme evaluation, and faculty development. In 2003 she was elected president of the Danish Association of Medical Education. She started on this PhD in Maastricht in 2000.